

Universidade de Lisboa
Faculdade de Ciências
Departamento de Biologia Vegetal



DISTRIBUIÇÃO E DIVERSIDADE DE ELEMENTOS GENÉTICOS
ASSOCIADOS À VIRULÊNCIA EM *STREPTOCOCCUS*
PNEUMONIAE

DISSERTAÇÃO

Joana Gomes Martins Lopes
MESTRADO EM MICROBIOLOGIA APLICADA
2013

Universidade de Lisboa
Faculdade de Ciências
Departamento de Biologia Vegetal



DISTRIBUIÇÃO E DIVERSIDADE DE ELEMENTOS GENÉTICOS
ASSOCIADOS À VIRULÊNCIA EM *STREPTOCOCCUS*
PNEUMONIAE

Dissertação orientada pelos Professores Doutores Mário Ramirez
(IM-FMUL) e Ricardo Dias (BioFIG-FCUL)

Joana Gomes Martins Lopes
MESTRADO EM MICROBIOLOGIA APLICADA

2013



DISTRIBUIÇÃO E DIVERSIDADE DE ELEMENTOS GENÉTICOS ASSOCIADOS À VIRULÊNCIA EM *STREPTOCOCCUS PNEUMONIAE*

Joana Gomes Martins Lopes

DISSERTAÇÃO DE MESTRADO

2013

Esta dissertação foi realizada na Unidade de Microbiologia Molecular do Instituto de Medicina Molecular (IMM) sob a orientação direta do Professor Doutor Mário Ramirez no âmbito do Mestrado em Microbiologia Aplicada da Faculdade de Ciências da Universidade de Lisboa.

O Professor Doutor Ricardo Dias foi o orientador interno designada no âmbito do Mestrado em Microbiologia Aplicada da Faculdade de Ciências da Universidade de Lisboa.

AGRADECIMENTOS

Em primeiro lugar quero agradecer ao Professor Doutor Mário Ramirez, meu orientador externo, pela orientação ao longo de todo este projeto e a confiança depositada. Agradeço igualmente ao Professor Doutor José Melo Cristino, Diretor do Instituto de Microbiologia da Faculdade de Medicina de Lisboa, a oportunidade de integrar esta equipa, que ajudou muito no meu enriquecimento tanto a nível profissional como pessoal.

Agradeço ao Professor Doutor Ricardo Dias, meu orientador interno, pela revisão da dissertação e pelos seus conselhos.

Um agradecimento muito especial à Andreia Horácio, do Instituto de Microbiologia, e à Ana Cohen, do “Center for Infectious and Inflammatory Diseases” da Universidade de Houston no Texas, que começaram comigo este projeto e apoiaram-me durante todo o desenvolvimento do mesmo tanto na execução técnica como na discussão do trabalho e na transmissão de conhecimentos.

A todos os meus colegas do Instituto de Microbiologia agradeço a amizade, a companhia e a ajuda que me deram especialmente à Joana Lopes, à Raquel Garcia, à Catarina Pato, ao Jorge Miranda e à Sandra Matos.

Por fim a todos os meus amigos e à minha família que sempre me incentivaram e apoiaram para que conseguisse chegar até ao fim e por acreditarem em mim.

RESUMO

Streptococcus pneumoniae é uma bactéria de Gram-positivo e um microrganismo comensal que coloniza assintomaticamente o aparelho respiratório superior (nasofaringe), contudo pode causar ocasionalmente infecção no hospedeiro com particular incidência nas crianças, idosos e doentes com condições crônicas debilitantes. Devido à sua grande variabilidade genética, resultante de altas taxas de transferência horizontal de genes, existe um desigual potencial patogénico entre estirpes. Entender como a variação do conteúdo genómico de *S. pneumoniae* se correlaciona com a evolução da doença grave tem sido o foco de vários estudos. O objetivo deste trabalho foi determinar os elementos genéticos associados à virulência em *Streptococcus pneumoniae*. Para tal, decidiu-se avaliar o potencial patogénico de oito regiões, denominadas regiões de diversidade (RD), responsáveis por mais de metade da diversidade genómica existente entre os pneumococos e com um potencial papel na virulência deste microrganismo. Foram analisadas estirpes de serotipo 1 e 3, que causaram infecções invasivas e não invasivas, com diferenças significativas entre si. Procedeu-se também ao estudo funcional de uma proteína de superfície (PspC) com um conhecido papel para a virulência do pneumococo.

Os dados obtidos para a presença das RDs nas estirpes de serotipo 1 permitiram confirmar a grande homogeneidade genética existente em serotipos raramente presentes em colonização, por oposição aos serotipos que são usualmente encontrados na mesma, como o serotipo 3, que apresentam uma alta diversidade genética. Demonstrou-se que a presença de alguns *loci* aparentam ser uma propriedade clonal de alguns complexos clonais (CC), relacionados com um dos tipos de doença e outros, relacionam-se com os dois tipos de doença. Concluiu-se que a RD1 é uma propriedade clonal das estirpes invasivas de CC180, CC260 e CC378, enquanto a RD2 uma propriedade clonal das estirpes não invasivas dos CC180 e CC260. Por outro lado o regulador transcricional da família MerR (RD2) parece ser uma propriedade clonal das estirpes não invasivas novamente do CC180, o sistema toxina-antitoxina PezAT (RD3/4) relacionou-se com as estirpes invasivas do CC180 e também do CC260 e a RD7 é uma propriedade das estirpes invasivas dos CC260, CC378 e CC458 e uma propriedade das estirpes não invasivas do CC180. Assim, existem regiões mais importantes para uma infecção mais invasiva e outras mais necessárias para a colonização do pneumococo. A diversidade dos próprios fatores de virulência, como a PspC, parece ser muito variável entre estirpes e nem sempre relacionada com o serotipo ou com o CC das estirpes.

Palavras-chave: *Streptococcus pneumoniae*, serotipo 1, serotipo 3, fatores de virulência, Regiões de Diversidade e Proteína de Superfície Pneumocócica C.

ABSTRACT

Streptococcus pneumoniae is a Gram positive bacterium and a commensal microorganism of the human upper respiratory tract (nasopharynx). However, it can also cause infections mainly in young children, the elderly and patients with chronic debilitating conditions. There is an unequal pathogenic potential between strains due to the high variability of the pneumococcal genome, which results from the high levels of horizontal gene transfer. Many studies have focused their objectives in understanding how the variation of the genome correlates with the evolution of severe disease. The objective of this work was to determine the *S. pneumoniae* genetic elements associated to virulence. For that we decided to evaluate the pathogenic potential of eight regions of diversity (RD), responsible for more than half of the genetic diversity present in the pneumococcus and relevant for the virulence potential of this microorganism. Strains isolated from invasive and non-invasive infections of serotypes 1 and 3, with significant differences between them, were analyzed. We also studied the structure of the pneumococcal surface protein C (PspC), which has a known roll in pneumococcus virulence.

The data on the presence of RDs in strains of serotype 1 allows us to confirm the great homogeneity present in serotypes rarely found in colonization, in opposition to serotypes that normally are found in it, such as serotype 3, which presents a high genetic diversity. We demonstrate that some *loci* appear to be clonal proprieties of some clonal complexes (CC), related with one of the types of disease, and others appear to be related with both types of disease. RD1 is a clonal propriety of invasive strains of CC180, CC260 and CC378 while RD2 is a clonal propriety of non-invasive strains of CC180 and CC260. On the other hand the MerR family transcriptional regulator (RD2) appears to be a clonal propriety of non-invasive strains belonging to CC180, the PezAT toxin-antitoxin system (RD3/4) was related to invasive strains of CC180 and CC260 at last the RD7 is a clonal propriety of invasive strains of CC260, CC378 and CC458 and also a propriety of the non-invasive strains of CC180. Therefore there are regions more necessary for invasion and others more necessary for colonization. The diversity of virulence factors, along with PspC allelic variability, confers variability amongst strains and it is not always related to the serotype or with the CC of the strains.

Key-words: *Streptococcus pneumoniae*, serotype 1, serotype 3, virulence factors, Regions of Diversity and Pneumococcal Surface Protein C.

1. INTRODUÇÃO.....	1
1.1 O agente infeccioso <i>Streptococcus pneumoniae</i>	1
1.2 Serotipos Capsulares	1
1.2.1 Serotipo 1	2
1.2.2 Serotipo 3	2
1.3 Genotipagem em pneumococos	3
1.3.1 MLST	3
1.3.2 PFGE.....	4
1.4 Vacinas anti-pneumocócicas.....	4
1.5 Fatores de virulência	5
1.5.1 Cápsula polisacarídica.....	5
1.5.2 Pneumolisina	6
1.5.3 Proteínas de ligação à colina	6
1.6 Regiões de diversidade.....	8
2. OBJETIVOS	10
3. MATERIAIS E MÉTODOS.....	11
<i>Capítulo I: Regiões Variáveis</i>	<i>11</i>
3.1 Coleção de estirpes	11
3.2 MLST	12
3.2.1 Obtenção do DNA cromossômico	12
3.2.2 Amplificação por PCR dos fragmentos internos de sete genes conservados.....	12
3.2.3 Purificação	13
3.2.4 Sequenciação.....	13
3.2.5 Identificação de alelos e perfis alélicos	13
3.3 Análise dos perfis de macrorestrição (PFGE)	13
3.4 Pesquisa das Regiões de Diversidade 1 a 8 e dos genes associados à virulência	14
3.4.1 Desenho dos oligonucleotídeos iniciadores	14
3.4.2 Otimização das condições de PCR	14
3.4.3 Obtenção do DNA cromossômico	14
3.4.4 Amplificação das RD e dos genes associados à virulência.....	14
3.5 Hibridação de ácidos nucleicos	15

3.6 Análise estatística	15
<i>Capítulo II: Proteína de superfície pneumocócica (PspC)</i>	<i>17</i>
3.7 Coleção de estirpes	17
3.8 Sequenciação da PspC	17
3.8.1 Análise das sequências	18
3.9 “Far-Western Blot”	18
3.9.1 SDS-PAGE e Transferência	18
3.9.2 Detecção.....	19
3.10 Análise informática	19
4. RESULTADOS.....	20
<i>Capítulo I: Regiões Variáveis</i>	<i>20</i>
4.1 População em estudo.....	20
4.1.1 Serotipo 1	20
4.1.2 Serotipo 3	21
4.2 MLST	21
4.3 Presença/ausência das Regiões de Diversidade.....	22
4.3.1 Serotipo 1	22
4.3.2 Serotipo 3	24
<i>Capítulo II: Proteína de superfície pneumocócica C.....</i>	<i>28</i>
4.3 Análise da estrutura secundária da PspC.....	28
4.4 Sequenciação da PspC.....	29
4.5 “Far-Western Blot”	31
5. DISCUSSÃO	33
5.1 Regiões Variáveis	33
5.2 Proteína de Superfície Pneumocócica C	37
5.2.1 Estrutura da PspC	38
5.2.2 Funções da PspC	39
5.3 Limitações e perspectivas futuras do estudo.....	41
6. CONCLUSÕES	42
7. ANEXOS	43
8. REFERÊNCIAS	46

Índice de Figuras

Figura 1: Comparação dos fenótipos dos serotipos pneumocócicos 1 e 3.	3
Figura 2: Fatores de virulência pneumocócicos.	5
Figura 3: Famílias das PspC.	7
Figura 4: “Primer Walking”	17
Figura 5: Distribuição das estirpes isoladas de infecção invasiva e não invasiva em função dos anos de 1999 a 2009.	20
Figura 6: Distribuição das estirpes isoladas de infecção invasiva e não invasiva em função do grupo etário e sexo dos doentes.	20
Figura 7: Distribuição das estirpes isoladas de serotipo 3 de infecção invasiva e não invasiva em função do CC.	21
Figura 8 A e B: Dendrograma simplificado construído com base nos perfis de macrorestrição, obtidos por PFGE.	22
Figura 9: Representação da associação estatística entre os 15 <i>loci</i> estudados e os CCs correspondentes.	26
Figura 10: Representação dos domínios A, B, R1, R2 e C da PspC3.4 e 7.1 obtidas através do programa Gor4..	28
Figura 11: Representação da ligação à membrana das PspC2.2, 3.4 e 3.7 e 10.1 e 11.1 obtidas através do programa PRED-TMBB.	29
Figura 12: Árvore “Neighbor-Joining” das PspCs.	30
Figura 13: Foto exemplificativa de Far-Western Blot	31
Figura 14: Alinhamento das sequências peptídicas da PspC..	32

Índice de Quadros

Quadro 1: Nomenclaturas e funções da PspC.....	7
Quadro 2: Componentes e respectivos volumes utilizados nas reações de PCR de MLST e RDs.....	12
Quadro 3: Programa de PCR utilizado para amplificar os fragmentos dos genes <i>aroE</i> , <i>gki</i> , <i>gdh</i> , <i>ddl</i> , <i>xpt</i> , <i>spi</i> e <i>recP</i> na técnica de MLST.....	13
Quadro 4: Programas de PCR para a pesquisa de RDs.....	16
Quadro 5: Componentes e respectivos volumes.....	18
Quadro 6: Programa de PCR utilizado para amplificar a <i>pspC</i>	18
Quadro 7: Presença dos <i>loci</i> nas estirpes de serotipo 1.	23
Quadro 8: Presença dos <i>loci</i> nas estirpes de serotipo 3.	24
Quadro 9: Associação entre o tipo de doença e a presença/ausência dos <i>loci</i>	26
Quadro 10: Distribuição dos <i>loci</i> nas estirpes invasivas (I) e não invasivas (NI) de serotipo 3.....	27
Quadro 11: Distribuição dos clones pelos tipos de doença	27
Quadro 12: Características estruturais de uma PspC representante de cada um dos 11 grupos de PspCs.	28
Quadro 13: Resumo dos resultados obtidos na sequenciação e no “Far-Western Blot” das PspCs.....	31
Quadro 14: Genes constituintes das RDs e sua localização em genomas já sequenciados.....	43
Quadro 15: Coleção de estirpes para a análise da PspC	43
Quadro 16: Oligonucleotídeos iniciadores utilizados nas reações de PCR	44
Quadro 17: Valores de “odds ratio” (OR) e de Fisher (valores p) referentes ao serotipo 1.	45
Quadro 18: Valores de “odds ratio” (OR) e de Fisher (valores p) referentes ao serotipo 3.	45

1. Introdução

1.1. O agente infeccioso *Streptococcus pneumoniae*

Streptococcus pneumoniae foi isolado pela primeira vez no final do século XIX e foi originalmente denominado de *Diplococcus pneumoniae*, devido à sua capacidade de formar pares (*Diplo*) de cocos (*coccus*) quando crescido em meio sólido e à sua propensão para se alojar nos pulmões provocando pneumonia. Apenas em 1974 passou a ser formalmente designado de *Streptococcus pneumoniae* (também referido como pneumococos), por existir uma grande semelhança ao nível do ácido desoxirribonucleico (DNA) com outras bactérias do género *Streptococcus*. Este microrganismo pertence ao grupo Mitis dos estreptococos que inclui 12 espécies filogeneticamente relacionadas através da sequência do ácido ribonucleico ribossomal (rRNA) 16S (57). É uma bactéria de Gram-positivo, alfa-hemolítica, solúvel em sais biliares e, geralmente, sensível à optoquina (2).

A colonização por pneumococos é assintomática, mas constitui o primeiro passo para o desenvolvimento de infeção. A partir da colonização podem desenvolver-se infeções que variam na sua gravidade, desde infeções não invasivas, tais como otite média aguda, a infeções invasivas, como meningite ou bacteriemia. Apesar de ser um microrganismo comensal colonizador do aparelho respiratório superior (nasofaringe), pode causar infeção no hospedeiro com particular incidência nas crianças, idosos e doentes com condições crónicas debilitantes (11, 54). Estima-se que colonize assintomaticamente a nasofaringe de 5 a 10% dos adultos saudáveis e 20 a 40% das crianças e jovens (81). As crianças são as principais portadoras e transmissoras de pneumococos na comunidade. Estudos de colonização feitos em Portugal revelaram que cerca de 61,3% das crianças entre seis meses a seis anos de idade, que se encontram em infantários, estão colonizadas por este microrganismo (83).

Considerando que a colonização precede sempre a infeção e que os valores da frequência de colonização são elevados, podemos considerar que a infeção invasiva é um evento relativamente raro. Contudo a Organização Mundial de Saúde (OMS) estima que ocorram em todo o mundo cerca de 14,5 milhões de episódios de infeção pneumocócica invasiva anuais em crianças menores de cinco anos de idade, dos quais, cerca de 826 000 resultam na morte dos doentes (74). A distribuição dos serotipos encontrados entre as estirpes responsáveis por infeção invasiva varia consideravelmente com inúmeros fatores tais como, a localização geográfica (42) e a idade do doente (93).

1.2 Serotipos Capsulares

A cápsula polisacarídica de *Streptococcus pneumoniae* apresenta um grupo diverso de polímeros com um papel essencial na virulência do organismo. É com base nos diferentes polissacáridos capsulares que se diferenciam os pneumococos em serotipos. Até à data já foram distinguidas serologicamente pelo menos 93

cápsulas diferentes (20), e foi determinada a estrutura de mais de metade. Os serotipos 1 e 3 são dos serotipos com maior relevância clínica em Portugal (1, 50). Estudos revelaram que o serotipo 3 foi o serotipo mais prevalente entre 1999 e 2008 (50), enquanto o serotipo 1 comporta-se como um importante agente patogénico primário (97).

1.2.1 Serotipo 1

O serotipo 1 continua a ser um dos serotipos pneumocócicos mais comuns desde a sua descrição em 1913 e tem sido repetidamente indicado como tendo um alto potencial invasivo, por raramente causar doenças não invasivas ou por colonizar assintomaticamente (18, 87). As estirpes de serotipo 1 causam frequentemente doença a indivíduos previamente saudáveis e comportam-se como um agente patogénico primário (97). Foi sugerido que este serotipo possui uma taxa de ataque invulgarmente alta por estar entre as causas mais comuns de doença invasiva, em muitas partes do mundo, e por ter uma elevada capacidade de causar surtos, apesar de raramente colonizar adultos ou crianças (18). Uma grande fração das infeções pneumocócicas invasivas (IPD- “Invasive Pneumococcal Diseases”) nas últimas décadas na Europa foi causada pelo serotipo 1 (32, 40), embora a sua taxa de prevalência seja muito mais reduzida nos Estados Unidos da América (EUA) (84).

Os perfis alélicos (“Sequence Types”, STs) deste serotipo foram agrupados em três grandes linhagens (A, B e C) num estudo de Brueggemann *et al.*, definidos como um grupo de estirpes que partilha quatro ou mais alelos. A linhagem A apresenta estirpes recolhidas sobretudo da Europa, EUA e Canadá, enquanto a maioria dos clones da linhagem B pertencem a estirpes de África e Israel e a linhagem C só possui o ST615 com estirpes recolhidas na América do Sul (18). Foi também considerado como tendo um potencial de doença invasiva muito homogéneo, ou seja, todos os clones deste serotipo estão distribuídos de igual forma na colonização e na doença invasiva (86).

Em Portugal, foi descrita uma associação entre IPD causada pelo serotipo 1 e crianças entre os seis e os 17 anos (1) e foi documentado que a proporção de estirpes de serotipo 1 isoladas de adultos com IPD diminui com a idade (50). No grupo pediátrico (0-17 anos) a prevalência do serotipo 1 aumenta com a idade, mas nos adultos (≥ 18 anos) a prevalência diminui. O número de casos de IPD aumentou após a introdução da vacina conjugada 7-valente (descrita no ponto 1.4) (50).

1.2.2 Serotipo 3

O serotipo 3 é comumente associado a doença invasiva, contudo é também o serotipo responsável pela maioria das otites originadas por pneumococos (79), sendo esta a doença não invasiva mais comum causada por *S. pneumoniae*. Este serotipo causa maioritariamente infeções em indivíduos com condicionantes e fatores de risco, pelo que foi descrito como um serotipo oportunista (97). Além disso, apesar de já ter sido associado a um baixo risco de IPD (17), o número de casos fatais de doentes infetados

com o serotipo 3 é elevado (97). Em modelos de ratinhos as estirpes de serotipo 3 são altamente virulentas (63).

Estas diferenças poderão ser explicadas por as estirpes de serotipo 3 apresentarem um potencial de doença invasiva heterogéneo (86), o que significa que existem clones com potenciais de doença invasiva distintos.

Em Portugal, a presença do serotipo 3 foi associada a idosos (50) e continua a ser o serotipo mais prevalente em adultos com IPD após a introdução da vacina conjugada 7-valente (descrita no ponto 1.4) (49). Também é um serotipo importante entre as crianças com IPD e é encontrado com regularidade em doentes com pneumonia complicada (49).

Em placa de gelose de sangue, a maioria das estirpes de serotipo 3 apresentam colónias mucoides características deste serotipo (Figura 1).

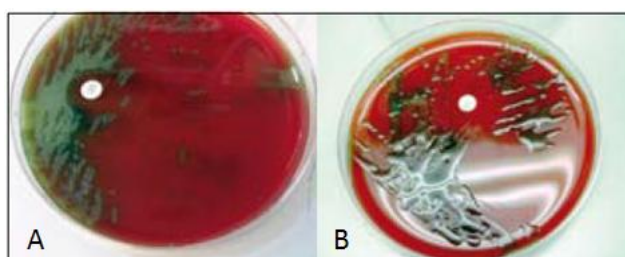


Figura 1: Comparação dos fenótipos dos serotipos pneumocócicos 1 e 3 (A e B, respetivamente). Disco impregnado com o antibiótico optoquina.

1.3 Genotipagem em pneumococos

Streptococcus pneumoniae é naturalmente transformável, podendo alterar profundamente o seu conteúdo genético, por esta razão é de extrema importância a utilização de técnicas que avaliem a diversidade genética existente na população pneumocócica. As duas técnicas correntemente mais utilizadas na genotipagem de pneumococos são: “Multi Locus Sequence Typing” (MLST) e eletroforese em gel de campo pulsado (“Pulsed Field Gel Electrophoresis”, PFGE).

1.3.1 MLST

A técnica de MLST consiste na comparação das sequências de DNA de fragmentos de sete genes conservados: *aroE* (chiquimato desidrogenase), *gdh* (desidrogenase da glucose-6-fosfato), *gki* (glucose cinase), *recP* (transcetolase), *spi* (peptidase sinal I), *xpt* (xantina fosforribosiltransferase) e *ddl* (D-alanina ligase). A cada gene é atribuído um número de alelo, de acordo com a sua semelhança com as sequências depositadas na base de dados (www.mlst.net). Ao conjunto formado pelos sete alelos obtidos é atribuído um número, denominado de ST. Ao conjunto formado pelo ST fundador e STs derivados do mesmo dá-se o nome de complexo clonal (CC). Estirpes que pertençam ao mesmo CC têm maior probabilidade de terem um ancestral comum próximo (33).

A comparação entre o perfil alélico das estirpes permite inferir sobre a sua semelhança genética. As relações entre as linhagens principais podem ser exploradas usando uma árvore filogenética construída a partir das sete sequências de MLST. Contudo, como nas árvores filogenéticas os eventos evolutivos mais recentes são

pouco representados, foi desenvolvido um procedimento útil baseado em STs relacionados (BURST). O algoritmo eBURST divide a população bacteriana em CC não sobrepostos e independentes e identifica um potencial fundador genotípico (ST fundador) de cada CC e os padrões da evolução descendente, mais prováveis, de todos os STs para o fundador (31). A inferência sobre a relação genética das estirpes é facilitada quando se utiliza um programa informático para análise, um exemplo deste tipo de programas é o “PHYLOVIZ” que utiliza este algoritmo. Os produtos da diversificação clonal são variantes do fundador que diferem em um dos sete *loci* examinados por MLST (“Single Locus Variant”, SLV). Subsequentemente, SLVs diversificam em DLVs (“Double Locus Variants”) e estes em TLVs (“Triple Locus Variants”), e assim sucessivamente.

MLST é uma técnica bastante reprodutível e não ambígua, que tem como principais vantagens permitir a comparação de resultados entre vários laboratórios e a criação de bases de dados internacionais, pelo que é uma importante ferramenta tanto na epidemiologia local, como global.

Em 1997 foi estabelecida uma rede de vigilância epidemiológica global, denominada de “Pneumococcal Molecular Epidemiology Network” (PMEN) (65), de modo a classificar e caracterizar os principais clones de pneumococos associados à resistência aos antibióticos. Atualmente, esta rede engloba não só os clones associados à resistência, mas também os mais relevantes em infecção invasiva que se encontram internacionalmente disseminados, tendo já sido descritos 43 clones (www.sph.emory.edu/PMEN/).

1.3.2 PFGE

A técnica de PFGE baseia-se na comparação dos perfis de fragmentos de DNA de diferentes estirpes, obtidos a partir da digestão do DNA total. A digestão ocorre através de uma endonuclease de corte pouco frequente que dá origem a fragmentos de DNA de elevada massa molecular, estes são posteriormente separados com recurso a uma variante da técnica de eletroforese em gel de agarose. Nesta variante as orientações do campo elétrico alteram-se periodicamente, permitindo a separação eficaz de fragmentos de DNA de elevado tamanho (90). Os padrões de bandas obtidos para cada estirpe são comparados com o intuito de inferir relações genéticas entre as estirpes (99). Muitos laboratórios fazem esta comparação com recurso a programas informáticos, para diminuir o erro, o tempo e as dificuldades associadas ao processo de análise manual (23).

A técnica de PFGE tem um elevado poder discriminatório, sendo bastante útil para a epidemiologia local. É considerada a técnica de eleição para a deteção de surtos (92).

1.4 Vacinas anti-pneumocócicas

O desenvolvimento de vacinas que previnam a infeção pneumocócica tornou-se fundamental devido à elevada incidência da mesma e ao aumento da resistência aos antibióticos. Atualmente existem dois tipos de vacinas anti-pneumocócicas disponíveis, as vacinas polissacarídicas e as vacinas polissacarídicas conjugadas.

Em Fevereiro de 2000 foi licenciada nos EUA a primeira vacina contra a infeção pneumocócica, a vacina conjugada 7-valente (PCV7), esta foi aconselhada a crianças menores de dois anos de idade. Na construção da vacina PCV7 escolheram-se os serotipos mais frequentes em infeção invasiva nos EUA que também eram aqueles que contabilizavam maior resistência aos antibióticos (serotipos 4, 6B, 9V, 14, 18C, 19F e 23F). Em Portugal a vacina PCV7 não faz parte do plano nacional de vacinação, contudo esteve disponível para venda pelo sector privado desde 2001 e em 2007 a sua cobertura era já de cerca de 70% (96). Após a introdução da PCV7 emergiram novos serotipos capsulares, que foram englobados em duas novas vacinas a vacina 10-valente (PCV10), em 2009 e a vacina 13-valente (PCV13), em 2010. A vacina PCV10 está disponível na Europa para imunização das crianças menores de dois anos de idade (106) e inclui, para além dos serotipos da vacina PCV7, os serotipos 1, 5 e 7F (89). Em Portugal, a vacina PCV13 encontra-se a substituir a vacina PCV7 desde Janeiro de 2010. Esta engloba todos os serotipos das vacinas PCV7 e PCV10 e mais três adicionais, os serotipos 3, 6A e 19A (109) e é aconselhada para a imunização de crianças com menos de dois anos, indivíduos de idade igual ou superior a 65 anos e de indivíduos com dois a 64 anos de idade que apresentem fatores de risco (ex. doenças crónicas e de imunossupressão) (6).

1.5 Fatores de virulência

Streptococcus pneumoniae possui um grande grupo de moléculas que podem ser consideradas fatores de virulência, definidos como substâncias produzidas pelo agente patogénico, envolvidas na iniciação e manutenção da infeção, que podem estar associadas à célula ou ser secretadas. A cápsula polissacarídica, essencial para a virulência (107), e a parede celular, que tem um papel essencial na inflamação induzida pelo organismo, são exemplos de fatores de virulência do pneumococo. Este possui ainda um grupo de fatores de virulência proteicos que incluem: proteínas de superfície, uma gama de enzimas e a toxina pneumolisina (Figura 2).

1.5.1 Cápsula polisacarídica

A cápsula polisacarídica é considerada o principal fator de virulência em pneumococos devido à sua função antifagocítica (51, 66, 67, 108). A presença de cápsula previne a interação do fator iC3b do complemento e do recetor Fc das imunoglobulinas, na superfície celular bacteriana, com os recetores da superfície das células fagocíticas, o que resulta na fuga ao sistema imunitário (72). A cápsula é crucial para a colonização, prevenindo a remoção mecânica pelo muco (73), restringindo a autólise e reduzindo a

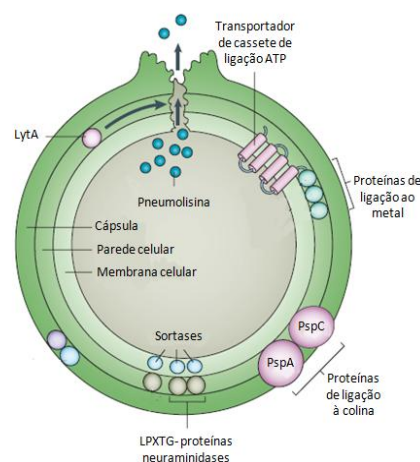


Figura 2: Fatores de virulência pneumocócicos.
Adaptado de (55)

exposição aos antibióticos (105). Estirpes sem cápsula são por norma avirulentas, ou causam infecções menos graves, como a conjuntivite (3).

1.5.2 Pneumolisina

A pneumolisina (PLY) pertence à família de toxinas que formam poros de espécies de bactérias de Gram-positivo. PLY tem a capacidade de lisar todas as células que contenham colesterol nas suas membranas e pode ter outros efeitos nas células. Por exemplo, foi demonstrado afetar a produção de mediadores pró-inflamatórios como o fator de necrose tumoral (TNF- α), a interleucina 1 β (IL-1 β) e a interleucina 6 (IL-6) (104). PLY é ainda capaz de ativar o sistema do complemento na presença e na ausência de anticorpos específicos (70, 77), esta capacidade é mediada pela ligação do fragmento Fc das imunoglobulinas.

1.5.3 Proteínas de ligação à colina

Streptococcus pneumoniae possui uma característica nutricional única, requer o catião colina, que é adquirido do meio e incorporado nas unidades repetidas de ácidos teicóicos e lipoteicóicos (5). Apesar de outras bactérias patogénicas conterem colina na sua parede celular, só o pneumococo expressa proteínas de superfície que se ligam especificamente a esta, permitindo a sua ligação à superfície celular. A função da maioria das proteínas de ligação à colina (“Choline Binding Proteins”, CBPs) é desconhecida, mas algumas foram estudadas e possuem um papel na patogénese desta espécie.

A autolisina LytA foi a primeira proteína a ser considerada como CBP, esta está implicada na patogenicidade por promover, como resultado da sua atividade de degradação da parede, a libertação de toxinas como a pneumolisina citoplasmática (8, 9) com a consequente danificação e destruição dos tecidos (22, 54, 76), e de componentes da parede nomeadamente, fragmentos de peptidoglicano e ácidos teicóicos (88, 102, 103), que desencadeiam fortes processos inflamatórios com um possível impacto negativo para o hospedeiro.

A primeira proteína de superfície exposta, que se descobriu ligar-se à colina, foi a proteína de superfície pneumocócica A (PspA) (110). PspA é estruturalmente organizada em 4 regiões distintas: uma região N-terminal funcional, uma região rica em prolina, uma região de ligação à colina e uma região C-terminal. A região N-terminal é constituída por 4 domínios com hélices α , que possuem uma região de ligação à lactoferrina (37). A ligação da PspA à lactoferrina bloqueia a acessibilidade do péptido, prevenindo a sua penetração na membrana bacteriana (91). PspA interfere também com o sistema do complemento (82) e é um importante fator de virulência, dado que, em modelos de sépsis, mutantes sem esta proteína tinham virulência reduzida (64).

A proteína de superfície pneumocócica C (PspC) possui um importante papel na patogénese de *S. pneumoniae*, contribuindo de diversas formas para a evasão ao sistema imunitário do hospedeiro (25, 27, 38, 53). Contudo a contribuição desta proteína para a virulência varia entre estirpes (56), uma vez que possui uma grande variabilidade genética. A nomenclatura PspC é apenas um nome genérico associado à

localização da proteína, consequentemente esta foi referida na literatura com diferentes designações associadas às diferentes funções biológicas estudadas. De forma a colmatar esta classificação dúbia foi proposta uma nova nomenclatura resumida no Quadro 1 (52). Recentemente foi descrita uma nova função desta proteína pela ligação à proteína de ligação C4BP, esta ligação parece ser dependente de uma variação específica da PspC, PspC4.4 (29). Deste modo, para evitar a lise bacteriana mediada pelo complemento, o pneumococo recruta a proteína C4BP, da mesma forma que recruta o fator H, reguladores centrais do complemento.

Quadro 1: Nomenclaturas e funções da PspC.

Sigla	Nome	Função	Referência	Nova classificação ¹	Estirpe	Serotipo
CbpA	“Choline-binding protein A”	Adesina	(85)	PspC3.1	R6	*
SpsA	“ <i>S. pneumoniae</i> secretory IgA binding protein”	Interage com a imunoglobulina A secretória	(38)	PspC2.1	ATCC33400	1
PbcA	“C3-binding protein A”	Ligação com o fator C3 do complemento	(25)	PspC3.1	CP1200	*
Hic	“Factor H-binding inhibitor of complement”	Ligação com o fator H do complemento	(53)	PspC11.4	A66	3

* Estirpes sem cápsula. ¹Classificação proposta por (52).

A nova classificação visa distinguir as variedades de PspC, organizando-as em 11 grupos diferentes de acordo com a sua estrutura e função (52) e em 2 grandes famílias, de acordo com o motivo de ligação da região C-terminal. A primeira família inclui as PspC dos grupos 1 a 6, que possuem uma região com repetições do motivo de ligação à colina na região C-terminal, que liga a proteína à superfície bacteriana através de interações com a colina presente nos ácidos teicóicos e lipoteicóicos. Enquanto a segunda inclui os grupos 7 a 11 que têm um motivo de ligação LPxTG típico de bactérias de Gram-positivo e que liga a proteína covalentemente ao peptidoglicano (Figura 3).



Figura 3: Famílias das PspC. A imagem **A** corresponde à família das proteínas PspC com a região C-terminal de ligação à colina (grupos 1 a 6). A imagem **B** corresponde à família das proteínas PspC com a região C-terminal com âncora LPxTG (grupos 7 a 11). Sequências homólogas estão desenhadas com caixas da mesma cor e alinhadas. As sequências sinal estão representadas a preto, as hélices α a amarelo, as sequências aleatórias em hélice (“random coil”) a cinzento, as regiões ricas em prolina a verde, a âncora de colina a azul e a âncora LPxTG a vermelho. Os asteriscos representam o motivo de ligação YPT. Adaptado de (52).

Apesar das diferenças entre as PspC, existe uma organização comum composta por um domínio N-terminal com hélices α expostas na superfície bacteriana, seguidas de uma região rica em prolinas e um motivo de

ligação à superfície celular (12). A sequência N-terminal exibe um grande número de repetições do motivo LZ (“Leucine Zipper”) (58) que aglomera cinco domínios denominados de A, B, R1, R2 e C (60). Apesar da falta de informação sobre a estrutura completa da PspC, a estrutura solúvel dos domínios R foi determinada usando espectroscopia de ressonância magnética (60). Entre as estruturas de hélices α foi localizado um hexapeptído responsável pela ligação à imunoglobulina A secretória e ao fator C3 do complemento. Este motivo varia mas tem sempre em comum a sequência YPT (Tirosina, Prolina, Treonina) que se supõe ser crítica para a ligação a estes fatores do complemento (39). Nas proteínas com ligação à colina existem dois motivos de ligação YPT, enquanto nas proteínas de ligação LPxTG este motivo só existe uma vez nas proteínas do grupo 7. O domínio C-terminal é responsável pela ligação à superfície celular e é formado por oito repetições do motivo de ligação à colina ou por um motivo de ligação LPxTG.

1.6 Regiões de diversidade

Como já foi referido esta é uma espécie naturalmente competente, transformável e a sua variabilidade genética é resultado de altas taxas de transferência horizontal de genes (“Horizontal Gene Transfer”, HGT) (46). Deste modo, a maioria das estirpes pneumocócicas pode adquirir DNA do ambiente e recombinar sequências com o seu cromossoma (35), resultando na substituição de fragmentos de DNA por sequências homólogas de outros genomas e na aquisição de novos genes de organismos dadores (30). As infeções causadas por pneumococos diferem muito na sua severidade e tal é atribuído em parte ao desigual potencial patogénico existente entre estirpes. Entender como a variação no conteúdo genómico de *S. pneumoniae* se correlaciona com a evolução da doença grave tem sido o foco de vários estudos.

Um estudo de Donati *et al.* analisou 44 estirpes de pneumococos e concluiu que apenas 48% do genoma era comum a todas as estirpes (30). *S. pneumoniae* possui, portanto, um grande repertório de genes não essenciais. Vários estudos têm avaliado esses genes não essenciais e se codificam fatores de virulência. Em 2001, através da técnica molecular CGH (“Comparative Genomic Hybridization”), determinaram-se 13 regiões, referidas como regiões de diversidade (RDs) por serem responsáveis por mais de metade da diversidade genómica existente entre os pneumococos (100). Desde essa data, já foram referenciadas 41 RDs (10, 16, 47, 75, 95). As RDs são muitas vezes compostas por conteúdo em GC atípico, são flanqueadas por sequências de inserção ou elementos genéticos, codificam genes semelhantes a fatores de virulência reconhecidos ou codificam genes hipotéticos ou com função desconhecida que foram demonstrados por STM (“Signature-tagged mutagenesis”) como sendo essenciais para a virulência em ratinhos (44, 59, 78). Um estudo recente de 2011 realizou a comparação do genoma de seis estirpes de serotipo 1 e identificou várias RDs (41), incluindo algumas previamente identificadas (10, 35, 75, 95, 100). Este estudo referencia oito RDs que só estiveram presentes em estirpes consideradas altamente virulentas através de modelos de infeção

em ratinhos (41). Neste trabalho destacamos estas 8 RDs, denominadas de RD1-8, pelo seu aparente papel na virulência uma vez que apenas foram encontradas em estirpes altamente virulentas.

RD1 consiste num bacteriófago inserido entre os genes *SPP_0028* e *SPP_0083* na estirpe de referência P1031. Esta região codifica um gene semelhante ao gene *pblB* (“Platelet-binding protein B”) e uma endolisina (amílase lítica), ambos necessários para a virulência de *Streptococcus mitis* em modelos animais de endocardite infecciosa (69). A sequência nucleotídica da endolisina partilha 80% de identidade com a autolisina *LytA*, descrita em 1.5.3. **RD2** codifica um regulador transcricional da família MerR, uma proteína da família MuT/Nudix e uma potencial permease dependente de sódio. Encontra-se em P1031 entre os genes *SPP_0747* e *SPP_0751*. MerR é um regulador transcricional identificado em *S. pneumoniae* como necessário para a prevenção do stress oxidativo gerado pelo óxido nítrico (NO) (98). **RDs 3 e 4** consistem na ilha patogénica *PPI-1* que contém o sistema toxina-antitoxina *PezAT* (TA), implicado na virulência (15), e uma região acessória. A região variável *PPI-1* possui o operão *piaABCD* e codifica um transportador do tipo ABC, necessário para a obtenção de ferro e para a virulência em modelos animais de pneumonia e septicémia (13, 14). Apesar deste transportador se encontrar presente em todas as estirpes, a região 3’ de *PPI-1* apresenta uma grande variabilidade entre estirpes (15). Estas regiões encontram-se compreendidas entre *SPP_1048* e *SPP_1069* na estirpe P1031. RD3 apresenta o sistema TA e RD4 o regulador transcricional *Rgg/GadR/MutR* e um potencial operão que codifica proteínas hipotéticas e enzimas metabólicas, como a 3HIBDH (“3-hidroxyisobutyrate dehydrogenase”). Contudo o gene *rgg/gadR/mutR* foi referido como não sendo funcional nas estirpes altamente virulentas, devido a mutações identificadas (48). **RD5** inclui uma potencial metaloproteínase de zinco D (*ZmpD*), cujo contributo para a virulência ainda não é conhecido. Contudo outras metaloproteínases de superfície, como *ZmpC*, foram associadas à virulência em pneumococos (21). Em P1031, esta região localiza-se entre *SPP_1140* e *SPP_1198*. **RD6** contém uma permease de alta afinidade para com ferro e chumbo, localizadas em P1031 entre *SPP_1340* e *SPP_1343*. O papel para a patogénese desta permease é ainda desconhecido, porém outros transportadores de iões metálicos, como *PsaA*, têm um importante papel na patogénese (101). **RD7** codifica um regulador transcricional da família *ArsR* e um transportador do tipo ABC-2. Os reguladores da família *ArsR* são muitas vezes suscetíveis a iões metálicos, e tal como os reguladores MerR, foram implicados na resistência a stresses ambientais (19). A região encontra-se entre *SPP_1779* e *SPP_1781* em P1031. **RD8** é constituída por 2 sub-regiões, entre *SPP_1350* e *SPP_1372* em P1031. A sub-região RD8b1 codifica um transportador do tipo ABC, uma proteína da família hidrólase glicosídica e uma liase N-acetil neuramindidase (NANA). A sub-região RD8b2 inclui um segundo transportador do tipo ABC. A proteína da família hidrolase glicosídica expressa-se quando a bactéria está no sangue e nos pulmões e o gene que a codifica possui apenas 50% de identidade entre estirpes virulentas e avirulentas (41).

2. Objetivos

A cápsula polisacarídica é considerada o principal fator de virulência em pneumococos, contudo estirpes que expressem o mesmo tipo capsular podem apresentar diferenças na sua capacidade invasiva ou na sua virulência, o que implica que existem outros fatores, além da cápsula, a contribuir para o potencial de invasão / virulência das estirpes.

O objetivo deste trabalho foi caracterizar os elementos genéticos variáveis associados à virulência em *Streptococcus pneumoniae*, especificamente a:

- Análise da distribuição de oito regiões de diversidade genética, consideradas importantes para a virulência em *S. pneumoniae*, em estirpes invasivas e não invasivas de serotipo 1 e 3 (capítulo I dos pontos 3 e 4);
- Determinação do tipo de PspC num subgrupo de estirpes de *S. pneumoniae* e relação do mesmo com a função e virulência da proteína (capítulo II dos pontos 3 e 4).

3. Materiais e Métodos

Capítulo I: Regiões Variáveis

3.1 Coleção de estirpes¹

As estirpes analisadas neste trabalho foram obtidas por um sistema de vigilância epidemiológica existente em Portugal desde 1999. O Estudo Viriato é um estudo prospetivo e multicêntrico de vigilância nacional da suscetibilidade aos antimicrobianos de bactérias frequentemente responsáveis por infeções do aparelho respiratório adquiridas na comunidade, este estudo procura contribuir para o conhecimento dos valores nacionais de resistência aos antimicrobianos mais prescritos em Portugal no tratamento empírico das infeções respiratórias, acompanhando ainda a evolução das resistências mais significativas em cada uma das espécies bacterianas incluídas. De acordo com este sistema é pedido a vários laboratórios de bacteriologia de todo o país para recolher e enviar todos os pneumococos responsáveis por infeções do aparelho respiratório adquiridas na comunidade para o laboratório coordenador desta investigação, o Instituto de Microbiologia da Faculdade de Medicina da Universidade de Lisboa (68). As estirpes foram recebidas de 49 laboratórios hospitalares localizados em várias regiões de Portugal, incluindo a ilha da Madeira. Nos casos em que foram enviadas em conjunto estirpes de duas amostras biológicas diferentes relativas ao mesmo doente, apenas uma foi considerada. As estirpes foram conservadas a -80°C em tubos de congelação contendo meio “Tryptone Soya Broth” (Oxoid, Hampshire, Inglaterra) e 15% de glicerol.

Foi estudada uma coleção de 545 estirpes de pneumococos que incluem estirpes recuperadas em Portugal durante o período de 1999 a 2009, escolhidas aleatoriamente, e 5 clones descritos na coleção PMEN (PMEN27, 28, 29, 31 e 40). A estirpe PMEN27 corresponde ao clone Sweden¹-27 de serotipo 1 e pertencente ao CC217, a estirpe PMEN28 corresponde ao clone Sweden¹-28 de serotipo 1 e do CC306, a estirpe PMEN29 corresponde ao clone USA¹-29 de serotipo 1 e do CC2296 e a estirpe PMEN31 corresponde ao clone Netherlands³-31 de serotipo 3 e do CC180. O estudo englobou 63 estirpes do serotipo 1 e 482 do serotipo 3, destas 203 estirpes foram isoladas de infeção não invasiva. Consideraram-se como infeções invasivas, as infeções em que o microrganismo foi isolado de locais habitualmente estéreis e infeções não invasivas quando o microrganismo foi isolado de locais que poderão não ser estéreis por terem uma abertura para o exterior, ou quando o processo de recolha da amostra envolveu um procedimento que a pode ter contaminado com a microbiota dos locais adjacentes.

As estirpes de referência TIGR4, R6, G54 e P1031 foram utilizadas como controlo. As bactérias foram cultivadas em “Tryptic Soya Agar” [(TSA)(Oxoid, Hampshire, Inglaterra)] suplementado com 5% (v/v) de sangue de carneiro desfibrinado (Probiológica, Belas, Portugal) e incubadas a 37°C.

¹ No presente trabalho a palavra *estirpe* é utilizada para referir cada microrganismo isolado num produto biológico.

3.2 MLST

Neste trabalho determinou-se o perfil de MLST de 384 estirpes de serotipo 3. As estirpes analisadas foram escolhidas aleatoriamente, de forma a englobar no mínimo 80% do número total de estirpes do serotipo 3, presentes na base de dados do laboratório, relativamente a cada ano estudado de 1999 a 2009.

3.2.1 Obtenção do DNA cromossômico

O método de obtenção de DNA consistiu na fervura das células bacterianas. Procedeu-se ao crescimento das células em “Brain Heart Infusion” [BHI (Difco Laboratories, Detroit, EUA)] e transferiram-se 9 µL de cada suspensão bacteriana para microtubos de 1,5 mL contendo 491 µL de água purificada pelo sistema mili-Q (Millipore, Billerica, EUA). As células foram fervidas durante 2 min e em seguida arrefecidas em gelo durante 5 min.

3.2.2 Amplificação por PCR dos fragmentos internos de sete genes conservados

As reações de PCR foram efetuadas utilizando a enzima polimerase de DNA Taq (Promega, Madison, EUA), tampão fornecido pelo fabricante, oligonucleotídeos iniciadores (Invitrogen, Pasley, Escócia), uma mistura de desoxiribonucleotídeos trifosfatados (dNTPs) (Fermentas, Vilnius, Lituânia), cloreto de magnésio (MgCl₂) (Promega, Madison, EUA) e água purificada pelo sistema Mili-Q. As reações de amplificação foram feitas com recurso a um termociclador da marca BioRad, modelo “MyCycler Thermal Cycler” (Munique, Alemanha).

Para determinar os perfis de MLST das estirpes selecionadas amplificaram-se os fragmentos internos dos genes *aroE*, *gdh*, *gki*, *recP*, *spi*, *xpt* e *ddl*. As sequências dos oligonucleotídeos iniciadores utilizados encontram-se descritas no Quadro 16 (Anexos) e os componentes e as respetivas quantidades utilizadas na mistura de reação estão indicados no Quadro 2. O programa de PCR utilizado está descrito no Quadro 3.

Quadro 2: Componentes e respetivos volumes utilizados nas reações de PCR de MLST e RDs.

Componentes	MLST (µL)	RDs (µL)
Polimerase de DNA Taq (5U/µL)	0,2	0,1
Tampão (10X)	10	5
dNTPs (10 mM)	1	0,5
MgCl ₂	4	2
Oligonucleotídeo iniciador F	2	1
Oligonucleotídeo iniciador R	2	1
DNA	20	10
Água	10,8	5,4
Volume Total	50	25

Quadro 3: Programa de PCR utilizado para amplificar os fragmentos dos genes *aroE*, *gki*, *gdh*, *ddl*, *xpt*, *spi* e *recP* na técnica de MLST.

Passo	Temperatura	Tempo	Nº de Ciclos
Desnaturação	95°C	5 min	1 x
Desnaturação	95°C	1 min	
Hibridação	48°C (58°C)	45 seg	4 x
Alongamento	72°C	1 min	
Desnaturação	94°C	1 min	
Hibridação	65°C	45 seg	29 x
Alongamento	72°C	1 min	
Terminação	72°C	5 min	1 x

A temperatura entre parêntesis do passo de hibridação foi utilizada apenas para a amplificação do gene *recP*.

3.2.3 Purificação

A purificação dos produtos de PCR foi realizada utilizando um sistema de purificação de DNA “High Pure PCR Product Purification Kit”, de acordo com as instruções do fabricante (Sigma-Aldrich, Steinheim, Alemanha).

3.2.4 Sequenciação

O DNA purificado foi enviado para a empresa MacroGen (Amsterdão, Holanda) para ser sequenciado. A sequenciação foi feita a partir dos oligonucleotídeos iniciadores universais M13F (5'-GTA AAA CGA CGG CCA GT-3') e M13R-pUC (5'-CAG GAA ACA GCT ATG AC-3').

3.2.5 Identificação de alelos e perfis alélicos

Para a identificação dos alelos e perfis alélicos de cada estirpe utilizou-se o programa informático BioNumerics, versão 6.5 (Applied-Maths, Sint-Martens-Latem, Bélgica), e a base de dados de MLST (www.mlst.net). Neste programa, as sequências obtidas foram alinhadas e comparadas com sequências padrão disponíveis na base de dados. Nos casos em que se obtiveram novos alelos ou perfis alélicos, submeteram-se as sequências ao responsável pela base de dados de MLST de *S. pneumoniae*, para que fossem atribuídos novos números de alelo e/ou perfis alélicos.

3.3 Análise dos perfis de macrorestrição (PFGE)

Os padrões genómicos das estirpes depositadas na base de dados do laboratório foram analisados e comparados através do programa informático BioNumerics. É de salientar que apenas os padrões genómicos das estirpes invasivas recolhidas em Portugal se encontram nesta base de dados e por isso, apenas estas estirpes foram analisadas. Construíram-se dendrogramas utilizando o coeficiente de Dice, com as definições de otimização e tolerância de 1,0 e 1,5%, respetivamente, e o método de agrupamento hierárquico “Unweighted Pair Group Method with Arithmetic means” (UPGMA). Consideraram-se como pertencentes à mesma linhagem de PFGE todas as estirpes que apresentaram uma semelhança igual ou superior a 80% (23).

3.4 Pesquisa das Regiões de Diversidade 1 a 8 e dos genes associados à virulência

3.4.1 Desenho dos oligonucleotídeos iniciadores

Foram desenhados oligonucleotídeos iniciadores de modo a amplificar oito regiões de diversidade (RDs). Os oligonucleotídeos iniciadores utilizados neste estudo encontram-se listados no Quadro 16 (Anexos). A comparação dos genomas e desenho dos iniciadores foi realizado com a ajuda do programa Vector NTI (Invitrogen). Para tal realizou-se uma pesquisa prévia de sequências de genomas anotados no GenBank: 670-6B (NC_014498.1), ATCC-700669 (NC_011900.1), G54 (NC_011072.1), Hungary 19A-6 (NC_010380.1), JJA (NC_012466.1), P1031 (NC_012467.1), CGSP14 (NC_010582.1), 70585 (NC_012468.1), R6 (NC_003098.1), TIGR4 (NC_003028.3) e Taiwan 19F-14 (NC_012469.1). As sequências foram alinhadas e os iniciadores foram desenhados nas zonas mais conservadas, de modo a amplificar as regiões completas 1, 2 e 7. Nas restantes não foi possível seguir esta estratégia pela diversidade das sequências e pelo seu tamanho. Deste modo escolheram-se genes, dentro das regiões, atendendo ao seu potencial papel para a virulência, identificadas no Quadro 14 (Anexos) e descritas em 1.6. Os iniciadores foram desenhados com a mesma metodologia das regiões. Para o gene codificante da proteína da família das hidrólases glicosídicas, por este possuir 50% de identidade entre estirpes virulentas e avirulentas, criaram-se 2 pares de iniciadores: um baseado na sequência da estirpe virulenta (P1031) e outro baseado na estirpe avirulenta (Hungary 19A-6). Todos os iniciadores foram desenhados de modo a apresentarem 19 a 25 pares de bases (bp-“base pairs”) de comprimento, regiões poucas ricas em (A+T) e (C+G) e não formarem dímeros ou grampos.

3.4.2 Otimização das condições de PCR

Para cada oligonucleotídeo iniciador foi calculada, empiricamente, a temperatura de hibridação pela fórmula: $T_m = [2(A+T) + 4(C+G)]$. As condições de PCR foram otimizadas de forma a realizar o maior número possível de reações multiplex, de modo a minimizar o número total de reações de PCR. Para isso foram testadas várias temperaturas de hibridação, de modo a agrupar as reações de PCR que tivessem tempos de alongamento iguais, temperaturas de hibridação semelhantes e tamanhos de fragmentos distintos, de modo a ser possível distinguir as bandas num gel de agarose. Foram realizadas cinco reações de PCR em multiplex para os *loci*: RD1 e Fe^{2+}/Pb^{2+} , ZmpD e Glycosídico_vir, Rgg/GadR/MutR e RD7, Int e ABC e Lyt e PblB.

3.4.3 Obtenção do DNA cromossómico

O DNA foi obtido através da fervura das células como descrito em 3.2.1

3.4.4 Amplificação das RD e dos genes associados à virulência

Apenas as estirpes que amplificaram o gene/região foram consideradas como portadoras das respetivas RD, exceto nas RDs 1 e 7. Nestas regiões os iniciadores foram desenhados nos genes flanqueantes. Devido ao tamanho das regiões (aproximadamente 3500-4000 bp, respetivamente) e ao facto de que a Taq polimerase poderá não ter eficiência para amplificar tal fragmento. Consideraram-se as estirpes como portadoras de ambas as regiões quando não há amplificação de qualquer fragmento.

Contrariamente, a amplificação de um fragmento de aproximadamente 300 bp é indicador de que as estirpes não apresentaram as regiões (o fragmento amplificado corresponde apenas a fragmentos dos genes flanqueantes às regiões 1 e 7). As reações de PCR foram realizadas como descritas em 3.2.2, os oligonucleotídeos iniciadores utilizados encontram-se listados no Quadro 16 (Anexos), os componentes e as respectivas quantidades utilizadas na mistura de reação estão indicados no Quadro 2 e os programas de PCR utilizado estão descritos no Quadro 4.

3.5 Hibridação de ácidos nucleicos

Os resultados negativos por PCR das regiões 1 e 7 foram confirmados pela técnica de “Dot blot”. Para as estirpes analisadas por “Dot blot” foi diluída uma cultura com uma densidade ótica (DO) de aproximadamente 2 em 800 µL de tampão de lise [0,4 M de NaOH + 10 mM de EDTA (“Ethylenediamine tetraacetic acid”)] e foram incubadas a 80°C durante 20 min. Para cada estirpe, foi transferida 50 µL desta preparação para uma membrana de náilon (Hybond N⁺; Amersham, Uppsala, Suécia) usando um sistema de vácuo (Pharmacia LKB Biotech, Uppsala, Suécia), e as membranas foram hibridadas com sondas específicas de DNA e detetadas por um sistema de quimiluminescência da Thermo Scientific (Massachusetts, EUA). As sondas foram geradas por PCR a partir de estirpes de referência, como descrito anteriormente, os produtos de PCR foram purificados utilizando um kit de purificação de produtos de PCR “High-Pure” (Roche, Mannheim, Alemanha) e marcadas com biotina de acordo com as instruções do fabricante pelo sistema “North2South” (Thermo Scientific, Massachusetts, EUA).

3.6 Análise estatística

Para analisar a diversidade de uma população utilizou-se o índice de diversidade de Simpson (SID), que indica a probabilidade de duas estirpes, escolhidas aleatoriamente de uma população, serem de tipos diferentes (24). Para comparar a presença de dois genes ou a associação entre um *loci* e o CC utilizou-se o coeficiente de “Adjusted Wallace” (AW), que calcula a probabilidade de duas estirpes possuírem o mesmo *loci*, ou de estirpes pertencentes ao mesmo CC possuírem o mesmo *loci*, corrigido para a frequência esperada em caso de independência (94). Para determinar a força da associação entre dois valores de dados binários calculou-se o “odds ratio” (OR). Utilizou-se o método exato de Fisher bilateral (p) para determinar significâncias estatísticas e corrigiram-se os valores para a taxa de descoberta de falsos positivos (“False Discovery Rate”, FDR) (7). Foram considerados valores significativos quando $p \leq 0,05$. O SID e o AW foram calculados em www.comparingpartitions.info. Os valores de OR, os intervalos de confiança a 95% (95%CI) e os valores p e FDR foram determinados utilizando as bibliotecas “epitools” e “multitest” do programa informático R.

Quadro 4: Programas de PCR para a pesquisa de RDs.

	RD1												RD2						RD3		
	RD1			PblB			Int			Lyt			RD2			MerR			pezAT		
Passo	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos
Desnaturação	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x
Desnaturação	95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg	
Hibridação	50°C	30 seg	30 x	58°C	30 seg	30 x	51,5°C	30 seg	30 x	54,8°C	30 seg	30 x	58°C	30 seg	30 x	50°C	30 seg	30 x	54,4°C	30 seg	30 x
Alongamento	72°C	1 min		72°C	1 min		72°C	2 min		72°C	2 min		72°C	2,5 min		72°C	2 min		72°C	3 min	
Terminação	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x

RD4						RD5			RD6			RD7						RD8					
rgg/gadR/mutR			3HIBDH			ZmpD			Fe ²⁺ /Pb ²⁺			RD7			ABC			Glicosídico_avir			Glicosídico_vir		
T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos	T	Tempo	Ciclos
94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x	94°C	4 min	1 x
95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg		95°C	30 seg	
58°C	30 seg	30 x	50°C	30 seg	30 x	54,4°C	30 seg	30 x	50°C	30 seg	30 x	58°C	30 seg	30 x	50°C	30 seg	30 x	54,4°C	30 seg	30 x	50°C	30 seg	30 x
72°C	2 min		72°C	1 min		72°C	2 min		72°C	1 min		72°C	2 min		72°C	1 min		72°C	2 min		72°C	30 seg	
72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x	72°C	10 min	1 x

“T”- Temperatura, Int-integrase, Lyt-endolisina.

Capítulo II: Proteína de superfície pneumocócica (*PspC*)

3.7 Coleção de estirpes

Foi analisada uma coleção de 19 estirpes de pneumococos recuperados em Portugal durante o período de 1999 a 2009, pertencentes ao mesmo sistema de vigilância epidemiológico descrito em 3.1. Analisou-se uma grande variedade de *PspCs*, que englobaram 11 serotipos diferentes: 3, 5, 6A, 6C, 7F, 8, 9V, 14, 18C, 19A, e 23F, recolhidas tanto de doença invasiva como não invasiva. No total foram estudadas sete estirpes invasivas, quatro estirpes não invasivas e sete estirpes da coleção PMEN (Quadro 15 - Anexos). A estirpe TIGR4 foi utilizada como controlo. As bactérias foram cultivadas como descrito em 3.1.

3.8 Sequenciação da *PspC*

Para determinar a estrutura do gene *pspC* realizou-se a sequenciação completa do mesmo, nas estirpes analisadas. Devido ao tamanho deste gene (ente 2,5 Kb e 3 Kb) a sequenciação foi realizada através da estratégia “primer walking”. Este método aplica-se a fragmentos cujo tamanho é muito elevado para serem sequenciados numa única reação. Deste modo realizou-se a mesma estratégia de um outro estudo, utilizando um par de oligonucleotídeos iniciador em genes conservados adjacentes à proteína: IF43 (5'-AAT GAG AAA CGA ATC CTT AGC AAT G-3') e IF30 (5'-AAG ATG AAG ATC GCC TAC GAA CAC-3') (52). A partir das sequências obtidas desenharam-se iniciadores em cada uma das cadeias até se obter a sobreposição dos fragmentos (Figura 4). Foram desenhados 45 oligonucleotídeos para a sequenciação. A amplificação do gene *pspC* foi realizada com uma mistura de polimerases de DNA para PCR longos, [“Expand Long Template PCR system” (Roche, Mannheim Alemanha)], um tampão adequado para a enzima utilizada fornecido pelo fabricante, oligonucleotídeos iniciadores (Invitrogen, Pasley, Escócia), uma mistura de dNTPs (Fermentas, Vilnius, Lituânia) e água purificada pelo sistema Mili-Q. Os componentes e as respetivas quantidades estão descritos no Quadro 5. Os programas de PCR utilizados estão descritos no Quadro 6.

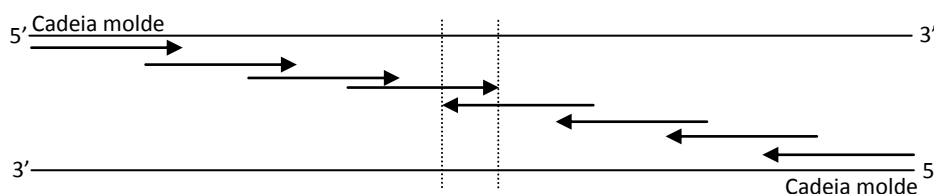


Figura 4: “Primer Walking”.

Quadro 5: Componentes e respectivos volumes utilizados nas reações de PCR da *pspC*.

Componentes	(μ L)
"Expand Long Template Enzyme mix" (5U/ μ L)	0,1
Tampão (10X)	5
dNTPs (10 mM)	1,75
Oligonucleotídeo iniciador F	1,5
Oligonucleotídeo iniciador R	1,5
DNA	10
Água	30,15
Volume Total	50

Quadro 6: Programa de PCR utilizado para amplificar a *pspC*.

Passo	Temperatura	Tempo	Ciclos
Desnaturação	94°C	2 min	1 x
Desnaturação	94°C	10 seg	10 x
Hibridação	48°C	30 seg	
Alongamento	68°C	8 min	
Desnaturação	94°C	15 seg	20 x
Hibridação	48°C	30 seg	
Alongamento	68°C	8 min + 20 seg ¹	
Terminação	68°C	7 min	1 x

¹Mais 20 seg em cada ciclo sucessivamente (x20). Por exemplo, o primeiro ciclo terá uma duração de 8 min, o segundo ciclo uma duração de 8 min e 20 seg e assim sucessivamente durante os 20 ciclos.

3.8.1 Análise das sequências

Os fragmentos obtidos foram montados pelo programa Vector NTI (Invitrogen) de modo a formar grupos, com fragmentos de sequências sobrepostas. A sequência consenso obtida foi traduzida para aminoácidos através da ferramenta on-line "EXPASy- Bioinformatics Resource Portal" (www.web.expasy.org/translate/). De seguida as sequências peptídicas obtidas foram alinhadas com as PspC já descritas pelo programa Geneious 6.1.6 (Biomatters, Ltd.) e analisou-se o grau de identidade entre elas.

3.9 "Far-Western Blot"

A ligação da PspC ao fator H foi analisada pela técnica "Far-Western Blot". Para tal começou-se por extrair as proteínas totais das estirpes analisadas. Neste método as estirpes são incubadas em 5 mL de THY, que contem "Todd Hewitt Broth" (Oxoid, Massachusetts, EUA) e 5 % de extrato de levedura ("Yeast Extract", Bacto, Pritchard, Austrália), até atingirem a fase estacionária, de seguida são transferidos 50 μ L de cultura para um novo meio até a solução atingir uma DO de 0,6. Depois são centrifugadas a 4000 rotações por minuto (rpm, Eppendorf, Modelo 5810R, rotor A-4-62, Nova Iorque, EUA) e ressuspensas em 2,5 mL de tampão PBS 1x ("Phosphate-buffered saline", contendo cloreto de cálcio e cloreto de magnésio, Gibco, Life Technologies Corporation), são novamente centrifugadas e ressuspensas em 100 μ L de Tris [Tris(hydroxymethyl)aminomethane, Bio-Rad Laboratories, Berkeley, Califórnia] 50mM a pH 7.5.

3.9.1 SDS-PAGE e Transferência

SDS-PAGE ("Sodium Dodecyl Sulfate-Polycrylamide Gel Electrophoresis") é um método utilizado para a separação das proteínas de acordo com o seu tamanho. O procedimento desta técnica envolve: 1) preparação de um gel de resolução a 12% de acrilamida e um gel de concentração a 5% de acrilamida e montagem do aparelho do gel, 2) mistura de 10 μ L de cada amostra com 2 μ L de SDS e aquecimento das soluções a 99°C durante 5 min seguido de arrefecimento em gelo durante mais 5 min, 3) carregamento das

amostras e corrida de eletroforese a 100V durante 2h15 e 4) transferência do gel de eletroforese para a membrana de nitrocelulose a 0,25 A durante 1h30.

3.9.2 Detecção

Após a transferência a membrana é corada com Ponceau S 1x (Sigma-Aldrich, St. Louis, EUA) e bloqueada com 5% de leite em PBS-T (PBS 1x com 0.1% de Tween 20, Sigma-Aldrich, St. Louis, EUA) durante 1h30. As proteínas são detetadas por ensaios de imunoblot. A membrana foi incubada com 1) o ligando (FH), 2) o anticorpo primário e 3) o anticorpo secundário. De seguida a existência de ligação é detetada pelo kit de deteção ECL (GE Healthcare, Uppsala, Suécia) de acordo com as indicações do fabricante. Utilizou-se o fator H humano a uma concentração de 0,04 µg/mL para 1/10000 mL de anticorpo primário de cabra anti-FH (Santa Cruz Biotechnology Inc., Heidelberg, Alemanha) e para 1/10000 mL de anticorpo secundário IgG anti-cabra (Santa Cruz Biotechnology Inc., Heidelberg, Alemanha).

3.10 Análise informática

De modo a melhor compreender a estrutura geral dos vários grupos de PspC analisou-se as sequências presentes no GenBank , até à data, através do programa on-line “GOR IV Secondary Structure Prediction Method” (www.npsa-pbil.ibcp.fr/cgi-bin/secpred_gor4.pl). O método GOR (Garnier-Osguthorpe-Robson) é um método informático, baseado na probabilidade de parâmetros derivados de estudos empíricos de proteínas com a estrutura terciária obtida por cristalografia de raio-X, para prever a estrutura secundária das proteínas. Ao contrário de outros métodos, o método GOR tem em conta não só a propensão de um aminoácido individual formar uma determinada estrutura secundária, mas também a probabilidade condicional do aminoácido formar a estrutura secundária dado os aminoácidos imediatamente adjacentes já formaram a estrutura. Deste modo, o método GOR analisa sequências de modo a prever se a estrutura secundária é formada por hélices α , folhas β ou sequências aleatórias em cada posição, baseando-se numa janela de 17 aminoácidos (34). De seguida, analisou-se as sequências com outro programa o PRED-TMBB (www.bioinformatics.biol.uoa.gr/PRED-TMBB/input.jsp). Este método baseia-se num modelo “Hidden Markov Model” que é capaz de prever as folhas β transmembranares de proteínas membranares. É fundamentado por uma base de dados não redundante que consiste em 16 proteínas da membrana externa com a sua estrutura conhecida através de resolução atómica (4). As estruturas obtidas foram comparadas entre si de modo a analisar as semelhanças existentes entre e dentro dos grupos.

4. Resultados

Capítulo I: Regiões Variáveis

4.1 População em estudo

No total foram analisadas 540 estirpes de serotipo 1 e 3 isoladas de infeção invasiva (n=337) e não invasiva (n=203), recolhidas entre 1999 a 2009 (Figura 5) e cinco estirpes da coleção PMEN. A Figura 6 representa a distribuição das estirpes isoladas de infeção invasiva e não invasiva por idade e sexo dos doentes, nos 10 anos estudados. A média de idade dos doentes foi de 50 anos de idade e observa-se que 60,2% (n=325) das estirpes pertencem a doentes com idades superiores a 50 anos. Relativamente ao sexo dos doentes, cerca de 65% (n=351) das estirpes foram isoladas de doentes do sexo masculino e 35% (n=189) de doentes do sexo feminino.

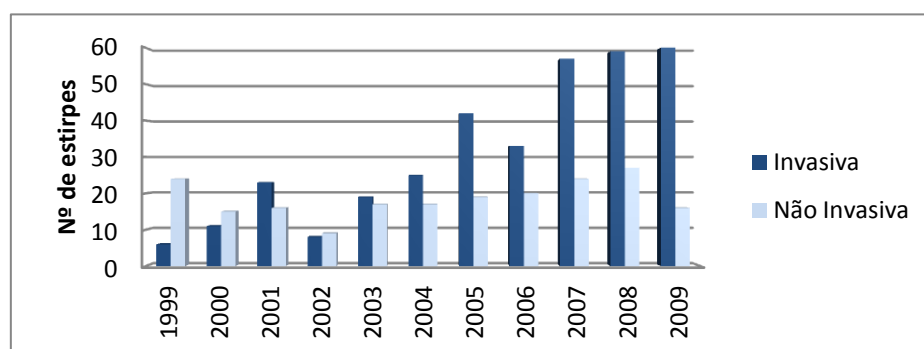


Figura 5: Distribuição das estirpes isoladas de infeção invasiva e não invasiva em função dos anos de 1999 a 2009.

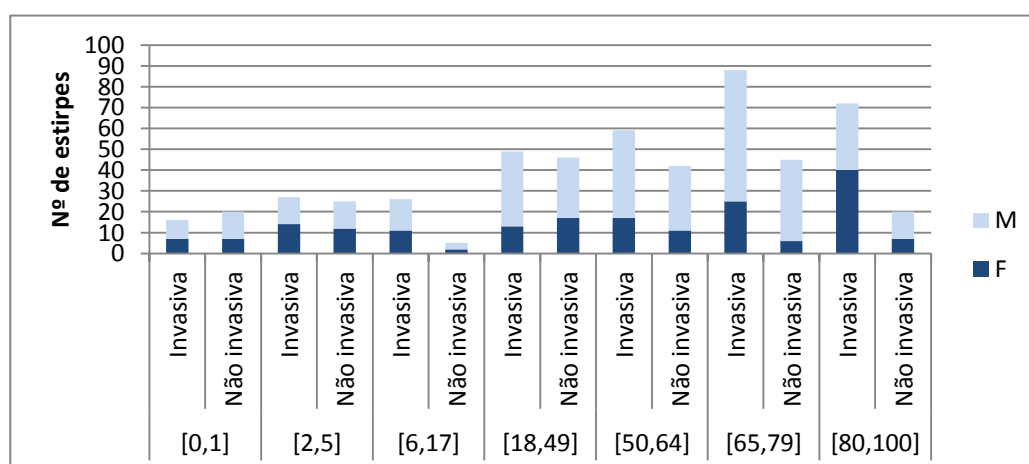


Figura 6: Distribuição das estirpes isoladas de infeção invasiva e não invasiva em função do grupo etário e sexo dos doentes. A idade dos doentes é apresentada em anos. "F"- sexo feminino e "M"- sexo masculino.

4.1.1 Serotipo 1

Foram analisadas no total 59 estirpes invasivas do serotipo 1, recolhidas em Portugal, e quatro clones da coleção PMEN. As estirpes de serotipo 1 recolhidas em Portugal distribuem-se por dois grupos de PFGE distintos (Figura 8A). Todas as estirpes de serotipo 1 recolhidas de infeção invasiva em Portugal fazem parte

do complexo clonal 306 (n=59). Consequentemente a diversidade dos perfis de MLST não é elevada, tal como indicado pelo índice de diversidade de Simpson [SID=0,692; (95%CI=0,598-0,785)]. Os perfis alélicos mais frequentes, correspondendo a cerca de 84% das estirpes de serotipo 1 foram os ST306 (n= 49,2%), ST304 (25,4%) e ST350 (9,5%), pertencentes ao CC306. As restantes estirpes (n=4) correspondem às estirpes PMEN27, 28, 29 e 40 que pertencem ao ST217, ST306, ST615 e ST306, respetivamente.

4.1.2 Serotipo 3

Para o serotipo 3 analisaram-se 481 estirpes, 57,8% (n=278) correspondem a estirpes recolhidas de infeção invasiva e os restantes 42,2% (n=203) a estirpes recolhidas de infeção não invasiva, em Portugal. As estirpes invasivas distribuem-se por apenas um grupo de PFGE (Figura 8B), as estirpes não invasivas não foram analisadas quanto aos seus perfis de PFGE. Juntou-se ainda à coleção uma estirpe PMEN31, correspondendo ao ST180. Também para este serotipo a diversidade dos perfis de MLST não é elevada, o índice de diversidade de Simpson é de 0,616 [95%CI= 0,573-0,660]. O CC180 foi o clone mais prevalente, em 59% (n=284) das estirpes, seguido do CC260 com 20% (n=98) de estirpes. Os três complexos clonais mais prevalentes na coleção (CC180, CC260 e CC378) perfazem um total de 94,5% (n=455) relativamente ao total de estirpes analisadas de serotipo 3. Apesar de terem sido estudadas uma maior percentagem de estirpes invasivas, não houve nenhum complexo clonal em que um dos tipos de infeção prevalecesse (Figura 7).

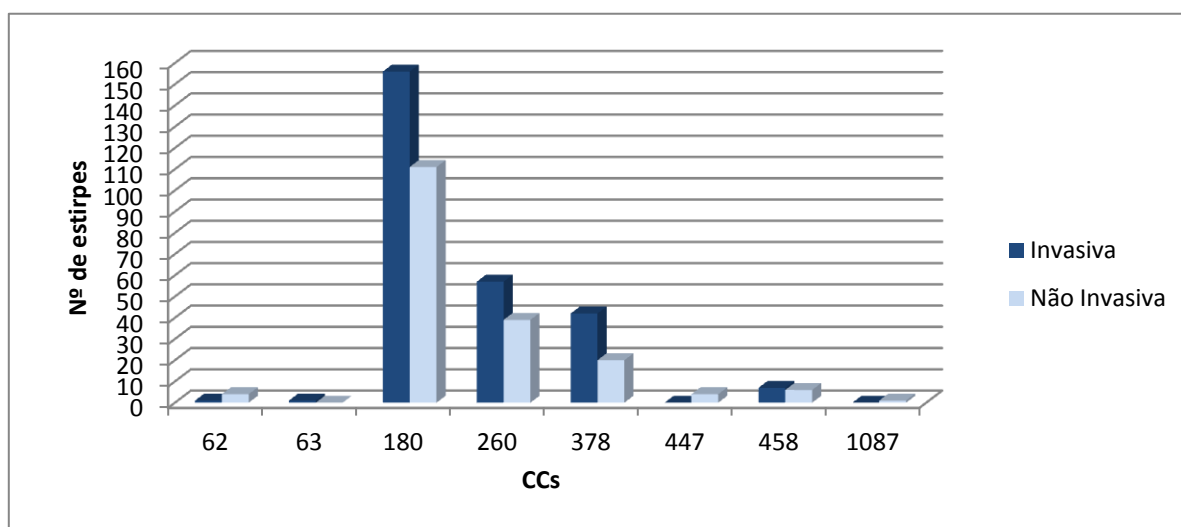


Figura 7: Distribuição das estirpes isoladas de serotipo 3 de infeção invasiva e não invasiva em função do CC.

4.2 MLST

Para avaliar a diversidade genética da coleção analisaram-se pela técnica de MLST 384 estirpes do serotipo 3, isoladas de infeção invasiva (n=185) e de infeção não invasiva (n=199), cujo ST ainda não tinha sido determinado. Após análise por MLST verificou-se a existência de novos alelos em 14 estirpes e de novas conjugações de perfis alélicos (ST) distintos em 10 estirpes. As 24 estirpes com novos STs distribuem-se por quatro CC, de acordo com a análise goeBURST.

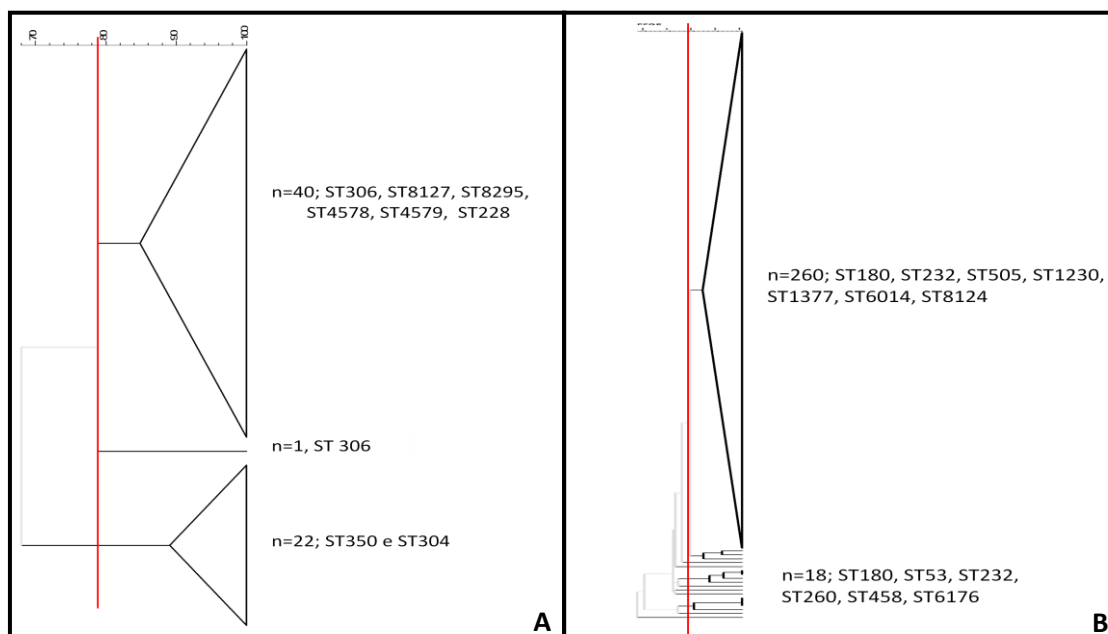


Figura 8 A e B: Dendrograma simplificado construído com base nos perfis de macrorestrição, obtidos por PFGE, das estirpes de serotipo 1 (Figura A) e das estirpes de serotipo 3 (Figura B), isoladas de infecção invasiva, entre 1999 e 2009. As estirpes com uma semelhança genética igual ou superior a 80% (indicada pela linha de corte vermelha) foram consideradas como pertencentes à mesma linhagem de PFGE. À frente de cada linhagem indica-se o número de estirpes detetadas e o ST correspondente.

4.3 Presença/ausência das Regiões de Diversidade

4.3.1 Serotipo 1

Os resultados obtidos para as estirpes de serotipo 1 quanto à presença das oito RDs estudadas, referidas em Harvey *et al.* (41), encontram-se resumidos no Quadro 7. Das estirpes de CC306 analisadas (n=61) apenas 12 estirpes possuem um dos *loci*, com exceção da RD8 onde 57 estirpes apresentaram o gene da família hidrólase glicosídica do tipo avirulento (*glyco_avir*). Contrastando com a estirpe representante do CC2296 que apresenta todos os genes exceto a endolisina *lyt*, o transportador do tipo ABC-2 da região 7 e o gene *glyco_avir*. A estirpe do CC217 apresenta os mesmo *loci* que o CC2296, exceto a RD2. Estes dados foram analisados estatisticamente pelo teste de Fisher e todos os valores foram corrigidos pelo teste de correção de falsos positivos (FDR) (Quadro 17 – Anexos). Os resultados são estatisticamente significativos para a ausência dos genes no CC306 contra os outros dois complexos para os *loci*: R1 completa, *PbIB*, integrase, *MerR*, sistema *pezAT*, *RGG/GadR/mutR*, *3HIBDH*, permease $\text{Fe}^{2+}/\text{Pb}^{2+}$ e RD7. Os dados são estatisticamente significativos para a presença dos genes em ambos os complexos CC217 e CC2296 para os *loci*: integrase, *MerR*, *RGG/GadR/mutR* e *3HIBDH*.

Para a presença da **RD1**, analisou-se a presença da região completa, do gene *pblB*, do gene codificante da integrase (*int*) e do gene codificante da endolisina (*lyt*). Nenhuma estirpe do CC360 apresentou a RD1 completa ou o gene *int*, contrariamente às estirpes dos restantes complexos que apresentaram estes *loci*. Quanto à procura do gene *pblB*, 11,4% (n=7) das estirpes do CC306 e 100% das estirpes do CC217 (n=1) e

CC2296 (n=1) apresentaram este gene. Para o gene *lyt* só 1,5% das estirpes do CC306 apresentou este gene, que corresponde a uma estirpe do ST304 (Quadro 7). A ausência/presença da RD1, PblB e integrase foram consideradas estatisticamente significativas para o CC306 versus CC217 e CC2296 (Quadro 17 - Anexos).

A amplificação do fragmento correspondente à **RD2** completa não foi possível em 98,3% (n=60) das estirpes de CC306 e na estirpe do CC217. Para os resultados da família MerR obteve-se uma percentagem de 100% para a ausência do gene no CC306 e 100% para a sua presença nos restantes CCs.

Nas **RD 3 e 4** pesquisou-se a presença dos genes codificantes do sistema *pezAT*, da família Rgg/Gadr/MutR e 3HIBDH. Nenhuma das estirpes do CC306 apresentou nenhum dos três *loci* estudados. Contrariamente ao CC2296 e CC217 que apresentaram todos os genes desta região.

Na **RD5** pesquisou-se a presença do gene *zmpD*, que esteve ausente em 96,7% (n=59) das estirpes de CC306 e presente em 100% das estirpes do CC217 e CC2296.

Para a **RD6** estudou-se a presença da permease de alta afinidade com Fe^{2+}/Pb^{2+} que esteve ausente em 100% das estirpes do CC306 e presente nas estirpes do CC2296 e CC217.

Só foi possível amplificar a **RD7** completa em uma estirpe do CC306 e em ambas as estirpes do CC217 e do CC2296. Nesta região também se estudou a presença do transportador do tipo ABC-2 que esteve ausente em todas as estirpes deste serotipo.

Como já referido, a **RD8** apresenta duas regiões, uma presente nas estirpes virulentas e outras nas estirpes avirulentas. Só uma estirpe do CC306 apresentou a região correspondente a estirpes virulentas, e as restantes apresentaram a região correspondente às estirpes avirulentas. As estirpes do CC217 e CC2296 apresentaram a região correspondente às estirpes avirulentas. Todas as estirpes apresentaram uma das regiões.

Quadro 7: Presença dos *loci* nas estirpes de serotipo 1.

ST	CC	RD1				RD2		RD3/4			RD5	RD6	RD7		RD8		Total
		RD1	PblB	Int	Lyt	RD2	MerR	pezAT	Rgg	3HIBDH	ZmpD	Fe/Pb	RD7	ABC	vir	avir	
227	306	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
228		0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
304		0	1	0	1	0	0	0	0	0	1	0	1	0	0	16	16
306		0	6	0	0	1	0	0	0	0	1	0	0	0	1	30	31
350		0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	6
4578		0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
4579		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
8127		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
8295		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Σ	306	0	7	0	1	1	0	0	0	0	2	0	1	0	1	59	61
217	217	1	1	1*	0	0	1*	1	1*	1*	1	1	1	0	1	0	1
615	2296	1	1	1*	0	1	1*	1*	1*	1*	1	1*	1*	0	1	0	1

Σ- Somatório das estirpes pertencentes ao CC306. *Int*-integrase, *Lyt*-endolisina, *vir*-família hidrólase glicosídica do tipo virulento, *avir*- família hidrólase glicosídica do tipo avirulento, *Rgg*- regulador transcricional Rgg/Gadr/MutR. Números a vermelho significam que os resultados são significativos segundo o teste de Fisher ($p \leq 0,05$). * Indica que é estatisticamente significativos pelo teste de correção de falsos positivos ($FDR \leq 0,05$).

4.3.2 Serotipo 3

Os resultados obtidos para o serotipo 3 respeitante à presença e/ou ausência das regiões de diversidade estudadas encontram-se agrupados quanto ao CC correspondente das estirpes no Quadro 8. Procedeu-se à análise estatística, semelhante à realizada para o serotipo 1, agrupando as estirpes por ST e por CC. Os resultados obtidos de ambas as formas foram semelhantes entre si, tendo-se optado por uma análise estatística por CC, de modo a simplificar a interpretação dos resultados (Quadro 18 – Anexos). Verifica-se que a maioria das estirpes pertencentes ao CC180 possui os genes de virulência pertencentes à RD1, contrariamente aos CC260, CC378 e CC458. Enquanto nas RDs 3/4 e 6 a maioria das estirpes do CC180 não apresenta os genes destas regiões, por oposição ao que acontece nas estirpes dos CC260, CC378 e CC458 que apresentam estes genes. Estes resultados são estatisticamente significativos pelo teste de Fisher ($p \leq 0,05$) para os *loci*: RD1, *PbIB*, integrase, *lyt*, RGG/GadR/MutR, permease Fe^{2+}/Pb^{2+} , RD7 e ABC.

Os restantes complexos clonais na sua maioria não apresentam resultados estatisticamente significativos, possivelmente devido ao baixo número de estirpes estudadas de cada clone. O CC62 apresentou resultados significativos quanto à presença dos genes *zmpD* ($p \leq 0,001$) e *glyco-vir* ($p \leq 0,001$) e quanto à ausência do fragmento correspondente ao sistema toxina-antitoxina *pezAT* ($p = 0,015$). O CC63 não apresentou resultados significativos, com exceção da presença do gene *3HIBDH* ($p = 0,015$) e *zmpD* ($p = 0,019$). Contudo só foi estudada uma estirpe deste complexo o que leva à pouca confiança nestes resultados, o mesmo aconteceu com o CC1087. Por último o CC447, apesar de ter um baixo número de estirpes ($n = 4$) apresentou resultados estatisticamente significativos para a presença do gene *rgg/gadR/mutR* ($p = 0,018$), ABC ($p = 0,018$) e *glyco_avir* ($p \leq 0,001$) e para a ausência da RD2 ($p = 0,018$) e RD7 ($p = 0,048$).

Quadro 8: Presença dos *loci* nas estirpes de serotipo 3.

CC	RD1				RD2		RD3/4			RD5	RD6	RD7		RD8		Total
	RD1	PbIB	Int	Lyt	RD2	MerR	pezAT	Rgg	3HIBDH	ZmpD	Fe/Pb	RD7	ABC	vir	avir	
62	1	2	2	2	3	5	2	3	3*	3*	0	3	3	3*	0	5
63	0	1	1	0	1	1	0	1	1	1*	0	0	0	0	0	1
180	56	223*	224*	186*	174	269	248	28	2	3	14	167*	16	7	7	284
260	32*	13	8	9	69	96	86	79*	0	0	86*	47	89*	2	0	98
378	15	12	5	4	44	71	67	62*	1	1	66*	33	64*	5	3	73
447	0	2	0	0	0	4	2	4*	0	0	3	0	4*	0	3*	4
458	0	0	1	0	12	15	15	0	0	0	0	5	0	0	0	15
1087	0	1	1	1	1	1	1	1	0	1*	0	0	1	0	0	1

Int-integrase, *Lyt*-endolisina, *vir*-família hidrólase glicosídica do tipo virulento, *avir*- família hidrólase glicosídica do tipo avirulento, *Rgg*- regulador transcricional RGG/GadR/MutR. Números a vermelho significam que os resultados são significativos segundo o teste de Fisher ($p \leq 0,05$). * Indica que é estatisticamente significativo pelo teste de correção de falsos positivos ($FDR \leq 0,05$).

Para a **RD1**, estudou-se a presença da região completa, do gene *pblB*, do gene *int* e do gene *lyt*. Apenas 20% ($n = 1$, $n = 56$, $n = 15$) das estirpes dos CC62, CC180 e CC378 e 32,7% ($n = 32$) das estirpes do CC260 apresentaram a RD1 completa. Quanto ao gene *pblB* 78,5% ($n = 223$) das estirpes do CC180 apresentou este gene, enquanto apenas 13,3% ($n = 13$) das estirpes do CC260, 16,5% ($n = 12$) das estirpes do CC378 e nenhuma das estirpes do

CC458 apresentaram este gene. Os resultados foram semelhantes aos restantes genes estudados desta região. O gene *int* está presente em 78,9% (n=224) e o gene *lyt* esteve presente em 65,5% (n=186) das estirpes do CC180, e ambos os genes estiveram ausentes em cerca de 90% (n=90; n=89) das estirpes do CC260, em 94,5% (n=68; n=69) das estirpes do CC378 e em 93,3% (n=14) e 100% (n=15) das estirpes do CC458, respetivamente. Nos restantes CCs os resultados não foram significativos.

Os resultados quanto à presença ou ausência de **RD2** não foram significativos para os três maiores clones analisados (CC180, 260, 378). Contudo, verifica-se que nestes complexos a percentagem de estirpes que possuem esta região completa (65%) é maior do que a do número de estirpes que não a possuem, ou possuem apenas parte da região (34%). Os únicos resultados significativos foram para a ausência desta região no CC447 (p=0,018). Os resultados da família MerR apresentaram uma percentagem de apenas 3,9% (n=19) de estirpes totais sem este gene. O valor é semelhante para os CC180, CC260 e CC378, os restantes não possuem nenhuma estirpe sem este gene. Este foi o *loci* estudado que esteve presente num maior número de estirpes (96%, n=462).

Nas **RD 3 e 4** pesquisou-se a presença dos genes codificantes do sistema *pezAT*, da família Rgg/Gadr/MutR e da enzima 3HIBDH. Os resultados quanto à presença do sistema toxina-antitoxina *pezAT* não apresentaram valores significativos, tendo a maioria das estirpes apresentado este sistema. O gene *rgg/gadR/mutR* apresentou resultados significativos para os CC180 (p≤0,001), CC260 (p≤0,001), CC378 (p≤0,001), CC447 (p=0,018) e CC458 (p=0,002). No CC180 este gene esteve ausente em 90% (n=256) das estirpes e no CC458 em 100% (n=15) das estirpes, contrariamente aos restantes complexos CC260, CC378 e CC447 onde este gene esteve presente em 80,4% (n=79), 84,9% (n=62), 100% (n=4), respetivamente. Apesar de o gene *3HIBDH* só apresentar resultados significativos para a sua presença nos CC62 e CC63 é de realçar que este esteve ausente em todas as estirpes dos CC260, CC447, CC458 e CC1087. No CC180 este gene estava ausente em 99,2% (n=282) e no CC378 em 98,6% (n=72) das estirpes.

Na **RD5** pesquisou-se a presença do gene *zmpD*, que esteve ausente em 97,9% (n=472) das estirpes totais. No entanto, 100% (n=1) das estirpes do CC63 e CC1087 e 60% (n=3) das estirpes de CC62 tinham este gene, apresentando resultados significativos.

Para a **RD6** estudou-se a presença da permease de alta afinidade com $\text{Fe}^{2+}/\text{Pb}^{2+}$, que apresentou resultados significativos para o CC180 e CC458 onde esteve ausente em 95% (n=270) e em 100% (n=15) das estirpes, por oposição ao CC260 e CC378 onde esteve presente em 87,8% (n=86) e 90,4% (n=66) das estirpes.

A amplificação completa de **RD7** foi possível em 58,8% (n=167) das estirpes de CC180. Os resultados não foram significativos para os restantes complexos exceto o CC447 que não amplificou o fragmento correspondente em 100% (n=4) das estirpes. O transportador do tipo ABC-2, analisado nesta região, esteve presente em 90% (n=89) das estirpes de CC260, em 87,7% (n=64) das estirpes de CC378 e em 100% (n=4)

das estirpes de CC447. Os resultados também foram significativos para o CC180 quanto à ausência do gene, estando este ausente em 94,3% (n=268) das estirpes.

Contrariamente ao que aconteceu para o serotipo 1, a amplificação das duas regiões de **RD8**, a presente nas estirpes virulentas e a presente nas estirpes avirulentas, não foi possível na maioria das estirpes. Apenas os CC62 e CC447 apresentaram resultados significativos: o CC62 apresentou o fragmento do gene virulento em 60% (n=3) das estirpes, enquanto o CC447 apresentou o fragmento do gene avirulento em 75% (n=3) das estirpes.

A associação entre a presença dos diferentes *loci* e os CCs foi analisada através do teste “Adjusted Wallace” (AW), como representado na Figura 9. Neste teste o coeficiente será de 1 se dois *loci* estiverem sempre presentes em simultâneo em diferentes estirpes de um dado CC, e de 0 nas situações inversas. Desta análise verificaram-se associações fortes entre os genes da RD1, sendo a mais alta de 0,753 [CI95%:0,677-0,830] entre os genes *int* e *pblB*. Os genes de RD3/4 não tiveram associações significativos exceto entre os *loci* *rgg/gadR/mutR* e Fe^{2+}/Pb^{2+} , porém esta não é muito elevada 0,604 [CI95%: 0,511-0,698]. Também se obteve uma associação elevada entre os genes *ABC* e Fe^{2+}/Pb^{2+} de 0,764 [CI95%:0,685-0,843], apesar de se localizarem em regiões diferentes. O *locus* que obteve uma maior associação com o CC foi novamente o gene Fe^{2+}/Pb^{2+} de 0,758 [CI95%: 0,668-0,847], seguido do gene *ABC* com 0,745 [CI95%:0,654-0,836].

Posteriormente tentou-se determinar se existiam diferenças na distribuição dos *loci* entre estirpes invasivas e não invasivas (Quadro 9).

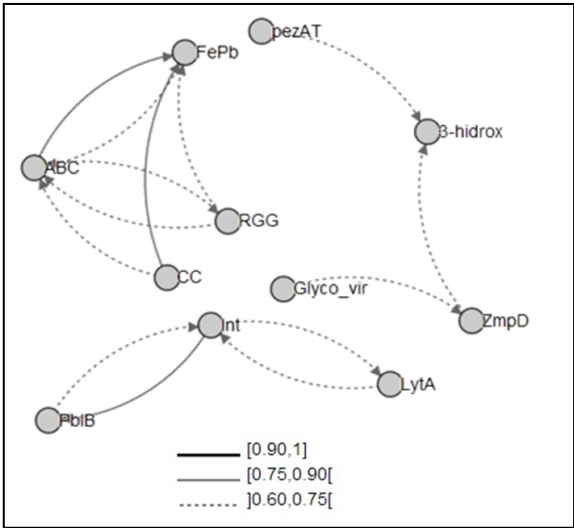


Figura 9: Representação da associação estatística entre os 15 *loci* estudados e os CCs correspondentes. As linhas correspondem aos coeficientes de AW > 0,6 de acordo com a legenda. Realizado na página eletrônica <http://darwin.phyloviz.net/ComparingPartitions/>

Quadro 9: Associação entre o tipo de doença e a presença/ausência dos *loci*.

	Invasivas		Não invasivas		OR	Valores de p	
	+	-	+	-		Fisher	FDR
RD1	98	189	6	197	17,892	≤0,001	≤0,001
PblB	148	130	106	97	1,034	0,854	0,915
Int	137	141	105	98	0,900	0,581	0,790
Lyt	124	154	78	125	1,298	0,190	0,408
RD2	217	61	87	116	4,704	≤0,001	≤0,001
MerR	261	17	201	2	0,152	0,004	0,011
pezAt	261	17	160	43	4,097	≤0,001	≤0,001
Rgg	106	172	72	131	1,127	0,567	0,790
3HIBDH	4	274	3	200	0,976	1,000	1,000
ZmpD	4	274	5	198	0,581	0,503	0,790
FePb	100	178	69	134	1,096	0,699	0,806
RD7	131	147	124	79	0,572	0,003	0,011
ABC	105	173	72	131	1,110	0,632	0,790
vir	13	265	4	199	2,446	0,137	0,342
avir	6	272	7	196	0,620	0,408	0,766

Int-integrase, *Lyt*-endolisina, *vir*-família hidrólase glicosídica do tipo virulento, *avir*- família hidrólase glicosídica do tipo avirulento, *Rgg*-regulador transcricional RGG/GadR/MutR. OR-“odds ratio”, FDR- teste de correção dos falsos positivos. OR > 1 indica uma associação entre as estirpes invasivas e os *loci*, OR ≤1 indica uma associação entre as estirpes não invasivas e os *loci*. Os valores a vermelho representam os valores estatisticamente significativos (p≤0,05).

Os resultados obtidos demonstram uma associação entre as estirpes invasivas e as RD1 e RD2 e com o sistema *pezAT*. Por outro lado, as estirpes não invasivas apresentam uma associação com o gene *merR* e com a RD7. É de realçar que existem associações com as regiões completas (RD1, RD2 e RD7), mas não com todos os seus genes constituintes. O caso que mais se destaca é a associação entre RD2 e as estirpes invasivas, contrária à associação do gene *merR*, localizado nesta região, com as estirpes não invasivas.

Uma vez que a presença/ausência de alguns *loci* aparenta ser uma propriedade clonal, procurou-se verificar se existe alguma associação de clones com o tipo de doença (Quadro 10). Os *loci* que apresentaram resultados estatisticamente significativos ($p \leq 0,05$) foram detetados nos CC180, CC260 e CC378, isto é, os mesmos que obtiveram resultados significativos na análise anterior, para a associação entre a presença dos *loci* e os CC e que são também os mais frequentes (Quadro 8).

Quadro 10: Distribuição dos loci nas estirpes invasivas (I) e não invasivas (NI) de serotipo 3.

	RDs	CC	62	63	180	260	378	447	458	1087
RD1	RD1	I	1	0	51*	31*	15*	0	0	0
		NI	0	0	5	1	0	0	0	0
	PblB	I	0	1	127	9	10	0	0	0
		NI	2	0	95	4	2	2	0	1
	Int	I	0	1	125	6	4	0	1	0
		NI	2	0	99	2	1	0	0	1
	Lyt	I	0	0	115	6	3	0	0	0
		NI	2	0	71	3	1	0	0	1
RD2	RD2	I	1	1	129*	48*	32	0	6	0
		NI	2	0	45	21	12	0	6	1
	MerR	I	1	1	150	57	45	0	7	0
		NI	4	0	119	39	26	4	8	1
RD3/4	pezAT	I	0	0	153*	57*	44	0	7	0
		NI	2	0	95	29	23	2	8	1
	Rgg	I	1	1	22	45	37	0	0	0
		NI	2	0	6	34	25	4	0	1
	3-hidroxi	I	1	1	1	0	1	0	0	0
		NI	2	0	1	0	0	0	0	0
RD5	ZmpD	I	1	1	1	0	1	0	0	0
		NI	2	0	2	0	0	0	0	1
RD6	Fe/Pb	I	0	0	7	52	41	0	0	0
		NI	0	0	7	34	25	3	0	0
RD7	RD7	I	1	0	55	40*	30*	0	5*	0
		NI	2	0	112*	7	3	0	0	0
	ABC	I	1	0	11	54	39	0	0	0
		NI	2	0	5	35	25	4	0	1
RD8	vir	I	1	0	5	2	5	0	0	0
		NI	2	0	2	0	0	0	0	0
	avir	I	0	0	5	0	1	0	0	0
		NI	4	0	119	39	24	1	8	1

Int-integrase, *Lyt*-endolisina, *vir*-família hidrólase glicosídica do tipo virulento, avir-família hidrólase glicosídica do tipo avirulento, *Rgg*- regulador transcricional RGG/GadR/MutR. Números a vermelho significam que os resultados são significativos segundo o teste de Fisher ($p \leq 0,05$). * Indica que é estatisticamente significativo pelo teste de correção de falsos positivos (FDR $\leq 0,05$).

De modo a descartar a hipótese da correlação encontrada ser apenas um reflexo do comportamento dos clones em relação aos tipos de doença, realizou-se também uma análise da distribuição dos clones pelo tipo de doença (Quadro 11). Como se pode observar não há valores estatisticamente significativos, o que demonstra que não existe uma predisposição de nenhum clone para causar um dos tipos de doença.

Quadro 11: Distribuição dos clones pelos tipos de doença

CC	I	NI	OR	Valor p	FDR
62	1	4	0,177	0,166	0,749
63	1	0	Inf	1,000	1,000
180	163	121	0,914	0,638	0,852
260	59	39	1,105	0,731	0,852
378	47	26	1,353	0,302	0,749
458	7	8	0,617	0,428	0,749
1087	0	1	0,000	0,417	0,749

OR—"odds ratio", Valor p- valores de p pelo teste de Fisher, FDR- teste de correção dos falsos positivos. "Inf"- infinito.

Capítulo II: Proteína de superfície pneumocócica C

4.3 Análise da estrutura secundária da PspC

As proteínas de superfície pneumocócica C possuem uma grande variedade genética e funcional. O trabalho realizado por Iannelli *et al.* revelou a existência de duas grandes famílias: as PspC que se ligam à superfície através de motivos repetidos de colina e as PspC que o fazem através do reconhecimento da sequência LPxTG por sortases (52). Para melhor compreender as diferenças e semelhanças entre os grupos, procedeu-se à análise da sequência peptídica das PspC depositadas no GenBank através dos programas GOR4 e PRED-TMBB. Foram analisadas 61 sequências peptídicas de PspC depositadas no GenBank.

Com o programa GOR4 foi possível identificar a sequência sinal de cada proteína e os domínios A, B, R1, R2 e C (60). Observou-se que a sequência sinal é idêntica em todos os grupos e que o domínio A é o mais conservado dos domínios, principalmente dentro da família das proteínas com ligação colina, exceto nas proteínas do grupo 5, que possuem domínios com um maior número de aminoácidos. O grupo 3 foi o que apresentou maior variabilidade dentro dos grupos. Entre as duas famílias de PspC foi possível identificar diferenças estruturais no que se refere aos motivos de ligação à imunoglobulina A secretória e ao fator C3 do complemento: a família de ligação à colina apresenta sempre um ou dois motivos (NYPT) localizados após o domínio R1 e R2, enquanto a família de ligação LPxTG contem apenas um motivo FRNYPT após o domínio R1 no grupo 7 (Quadro 12). O domínio C também é diferente na família LPxTG, apresentando um maior número de aminoácidos ou estando fragmentado antes e após o motivo LPxTG, contrariamente às PspC da família de ligação à colina que possuem um único domínio C curto e coeso (Figura 10).

Quadro 12: Características estruturais de uma PspC representante de cada um dos 11 grupos de PspCs.

PspC	Família	Grupo	Motivo	Localização (aa)	Motivo	Localização (aa)	Tamanho Total (aa)
1.1	Colina	1	RRNYPT	167-172	RRNYPT	413-418	929
2.2	Colina	2	RRNYPT	202-207	RRNYPT	355-360	693
3.4	Colina	3	RRNYPT	202-207	RRNYPT	355-360	693
4.2	Colina	4	HRNYPT	144-149	—	—	866
5.1	Colina	5	FRNYPT	169-174	RRNYPT	420-426	869
6.1	Colina	6	HRNYPT	159-164	RRNYPT	321-326	681
7.1	LPxTG	7	YRNYPT	200-206	LPSTG	738-742	769
8.1	LPxTG	8	—	—	LPSTG	472-476	503
9.1	LPxTG	9	—	—	LPSTG	553-557	584
10.1	LPxTG	10	—	—	LPSTG	732-736	763
11.1	LPxTG	11	—	—	LPSTG	581-585	612

O estudo foi realizado pela análise das sequências peptídicas. Os motivos correspondem a sequências de aminoácidos, assim como a localização e o tamanho total correspondem à posição dos aminoácidos na sequência.

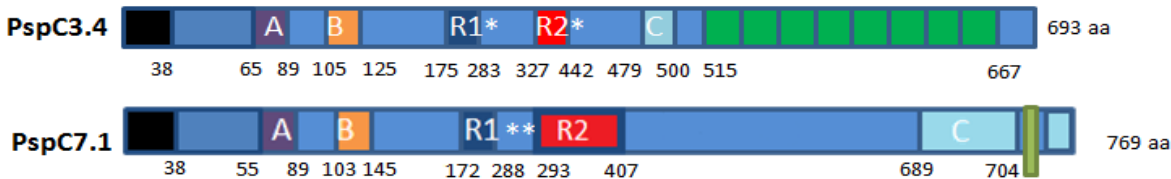


Figura 10: Representação dos domínios A, B, R1, R2 e C da PspC3.4 (família colina) e 7.1 (família LPxTG) obtidas através do programa Gor4. Os números representados em baixo dos domínios referem-se à localização dos aminoácidos na sequência. O retângulo preto representa a sequência sinal. O asterisco representa o motivo RRNYPT e dois asteriscos representam o motivo FRNYPT. Os retângulos verdes representam as repetições do domínio colina, e o retângulo verde claro representa o motivo LPxTG.

O programa PRED-TMBB prevê e representa os domínios transmembranares existentes em proteínas membranares, tendo-se obtido as imagens ilustradas na Figura 11. Verificou-se uma grande semelhança entre as estruturas das proteínas dos grupos da família LPxTG e uma maior diversidade entre as proteínas da família de ligação à colina, com diferenças mais acentuadas entre os elementos do grupo 3. A Figura 11 mostra um exemplo dessa diversidade entre as PspC3.4 e PspC3.7. Verificou-se também que a estrutura da proteína 2.2 e 3.4 é muito semelhante, corroborando os resultados obtidos pelo programa Gor4. Após alinhamento no programa Geneious 6.16 verificou-se uma identidade de 99,7%, correspondente apenas a diferenças em dois aminoácidos. Este valor é muito elevado, principalmente tendo em conta que foi superior ao obtido quando se alinhou sequências do grupo 3, por exemplo a identidade entre a PspC3.4 e PspC3.1 é de apenas 73,6%.

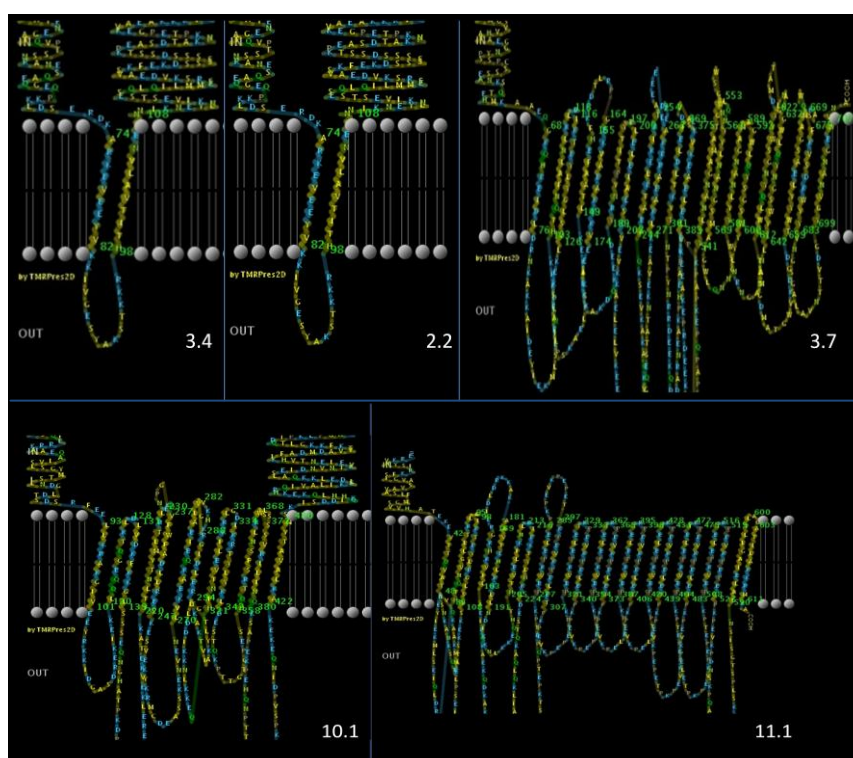


Figura 11: Representação da ligação à membrana das PspC2.2, 3.4 e 3.7 (família colina) e 10.1 e 11.1 (família LPxTG) obtidas através do programa PRED-TMBB.

4.4 Sequenciação da PspC

De modo a estudar-se a variabilidade da PspC em *S. pneumoniae*, sequenciaram-se 19 estirpes de 11 serotipos (3, 5, 6A, 6C, 7F, 8, 9V, 14, 18C, 19A e 23F) por “Primer Walking”. Das 19 estirpes analisadas, 15 aparentam ter uma única cópia do gene *pspC*, enquanto 3 possuem duas cópias diferentes do gene. O grupo de PspC foi determinado tendo por base a identidade entre as sequências peptídicas das PspC conhecidas e as sequências obtidas nas novas estirpes. A percentagem de identidade entre as estirpes foi variável (entre 22% a 84,3%), consequentemente a confiança no grupo de PspC determinado é limitada em muitas estirpes (Quadro 13). Das 22 sequências de *pspC* obtidas, seis apresentaram âncora LPxTG, três delas na segunda cópia do gene *pspC* em conjunto com uma cópia do gene com motivos de colina. As estirpes que possuem

duas cópias da PspC englobaram os grupos 4-10, 6-7 e 6-9 (Quadro 13). A sequência da PspC da estirpe PMEN39 da família de ligação à colina não alinhou com nenhuma PspC conhecida, sugerindo a presença de um novo grupo destas proteínas. Foram encontradas PspCs de todos os grupos, exceto do grupo 2. O grupo de PspC com maior predominância nesta coleção foram os grupos 3 (n=4) e 6 (n=5). Nesta coleção, as estirpes do serotipo 19A (n=4) apresentaram proteínas de ligação à colina ou duas PspC, enquanto as estirpes de serotipo 3 (n=2) possuem proteínas com âncora LPxTG.

A Figura 12 representa uma árvore sem raiz de todas as PspC analisadas, onde se pode visualizar uma clara separação entre as proteínas de ligação à colina e as proteínas de ligação LPxTG (linha a tracejado), sendo o grupo 7 o mais próximo das PspCs da família de ligação à colina.

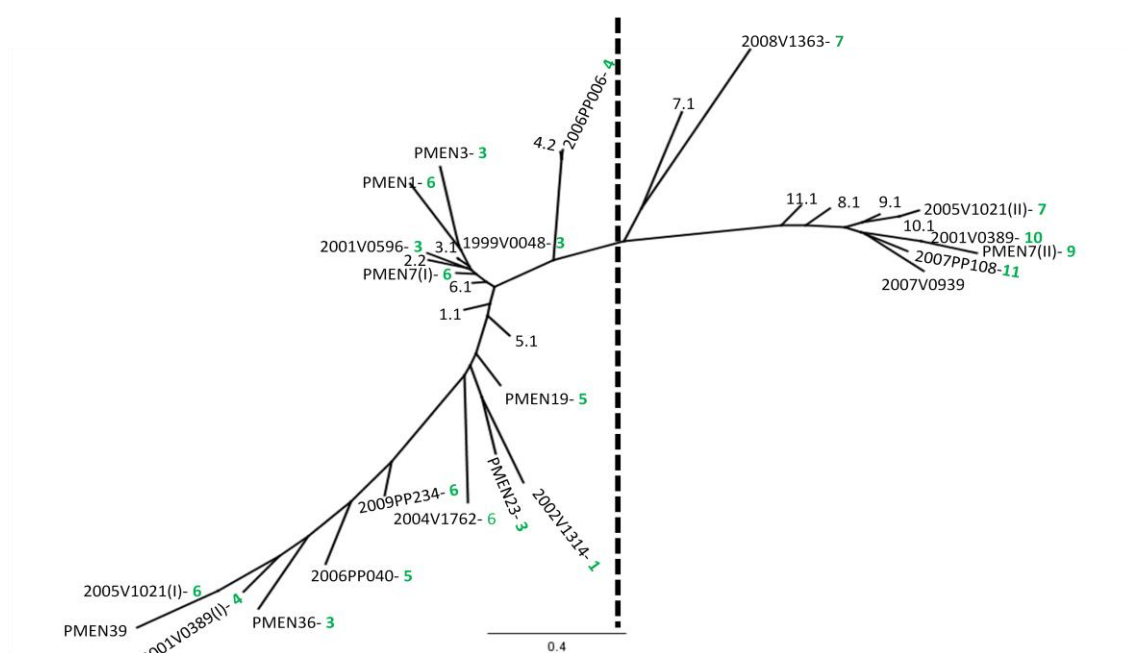


Figura 12: Árvore “Neighbor-Joining” das PspCs. A árvore é um filograma sem raiz gerado pelo método “Neighbor-Joining” com média de distâncias características e com matriz de score de aminoácidos Blosum62, construída no programa Geneious 6.16. O comprimento dos ramos entre as proteínas é proporcional à semelhança entre as sequências. Os números a verde indicam o grupo de PspC sugerido para as proteínas estudadas. A linha a tracejado separa as PspC da família com ligação à colina da família de PspC com ligação LPxTG.

Na Figura 14 observa-se o alinhamento de sequências peptídicas dos 11 grupos de PspC já descritas com as PspCs estudadas nesta coleção. Observou-se que todas as PspC possuem uma sequência sinal e que esta tem uma maior semelhança entre as estirpes de ligação à colina do que entre as estirpes com ligação LPxTG. Também se verificou a presença do hexapeptídeo de ligação à imunoglobulina A secretória e ao fator C3 do complemento, com a sequência YPT comum às estirpes da família de ligação à colina. Das 15 estirpes sequenciadas com motivos de colina, apenas cinco não possuem nenhum motivo YPT. Enquanto das seis estirpes sequenciadas com motivo LPxTG, apenas uma apresentou um motivo YPT. A estirpe com motivo YPT da família de ligação LPxTG apresentou uma maior identidade com as proteínas do grupo 7.

4.5 “Far-Western Blot”

Realizou-se “Far-Western Blot” de modo a determinar que estirpes se ligam ao fator H, além disso, pelo número de bandas foi possível averiguar o número de proteínas PspC presente em cada estirpe. Das 19 estirpes analisadas, quatro estirpes apresentaram duas bandas e as restantes só apresentaram uma banda. Os tamanhos das bandas variaram de 50 kDa a 100 kDa. Houve concordância em todas as estirpes no número de PspC existentes detetado por sequenciação e por “Far-Western Blot”, exceto em três estirpes: 2002V1314, 2008V1363 e PMEN7

(Quadro 13). As estirpes 2002V1314 e 2008V1363 apresentaram duas bandas, a primeira apresentou uma banda de 100 kDa e uma banda mais fraca na ordem dos 75 kDa, enquanto a segunda estirpe apresentou uma banda de 100 kDa e outra banda muito fraca de 50 kDa, porém devido ao tamanho desta, não deve ser referente à PspC (Figura 13).

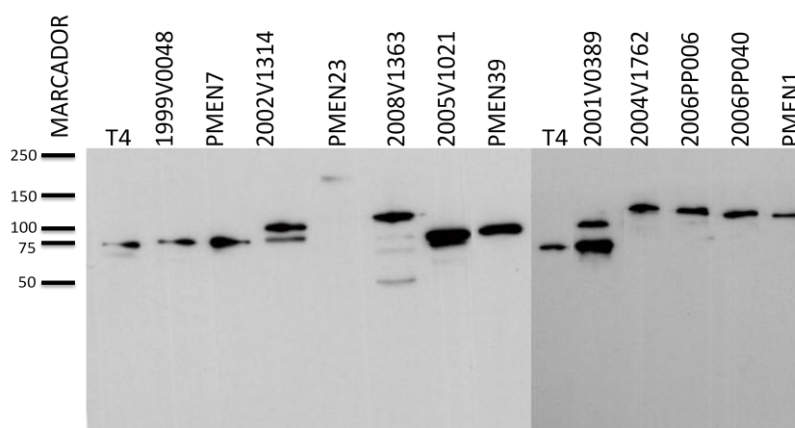


Figura 13: Foto exemplificativa de Far-Western Blot

Quadro 13: Resumo dos resultados obtidos na sequenciação e no “Far-Western Blot” das PspCs.

Serotipo	Sequenciação								“Far-Western Blot”	
	Estirpes	PspC	Alinhamento	Motivos	Localização	Motivos	Localização	Tamanho (aa)	FH	
									nº bandas	Tamanho (kDa)
4	T4	3.4	Colina	RRNYPT	202-207	RRNYPT	355-360	693	1	75
14	1999V0048	3 (84,3%)	Colina	LRNYPT	164-169	RRNYPT	327-330	784	1	75
19A	2001V0389	4 (22,8%) 10 (80%)	Colina LPxTG	—	—	— LPSTG	676-680	508 712	2	75 100
8	2001V0596	3 (78,3%)	Colina	FRNYPT	130-135	HRNYPT	283-288	680	1	50
6A	2002V1314	1 (60,3%)	Colina	RRNYPT	167-172	—	—	703	2	75 100
19A	2004V1762	6 (41,1%)	Colina	—	—	RRNYPT	337-342	540	1	100
6C	2005V1021	6 (21,9%) 7 (44 %)	Colina LPxTG	—	—	— LPSTG	358-362	663 389	2	70 75
14	2006PP0006	4 (76,8%)	Colina	RRNYPT	156-161	—	—	940	1	100
19A	2006PP0040	5 (22,5%)	Colina	—	—	—	—	715	1	100
3	2007PP108	11 (76,6%)	LPxTG	—	—	LPSTG	295-299	324	1	75
3	2007V0939	8 (69,6%)	LPxTG	—	—	LPSTG	—	349	1	75
9V	2008V1363	7 (37,5%)	LPxTG	YRNYPT	248-253	LPSTG	757-761	788	2	50 100
7F	2009PP234	6 (43,1%)	Colina	HRNYPT	165-170	—	—	680	1	75
23F	PMEN1	6 (49%)	Colina	—	—	—	—	823	1	100
9V	PMEN 3	3 (64%)	Colina	LRNYPT	164-169	RRNYPT	325-330	664	1	75
19A	PMEN7	6 (79,9%) 9 (51%)	Colina LPxTG	RRNYPT —	167-172	RRNYPT LPSTG	331-336 339-403	688 435	1	75
5	PMEN 19	5 (74,2%)	Colina	HRNYPT	120-125	—	—	314	1	75
6A	PMEN23	3 (44,9%)	Colina	RRNYPT	162-167	—	—	442	1	150
18C	PMEN 36	3 (59%)	Colina	—	—	—	—	773	1	75
7F	PMEN39	—	Colina	—	—	—	—	608	1	100

Os valores entre parêntesis na coluna da PspC referem-se à percentagem de identidade obtida entre a sequência de PspC a determinar e uma PspC do grupo correspondente.

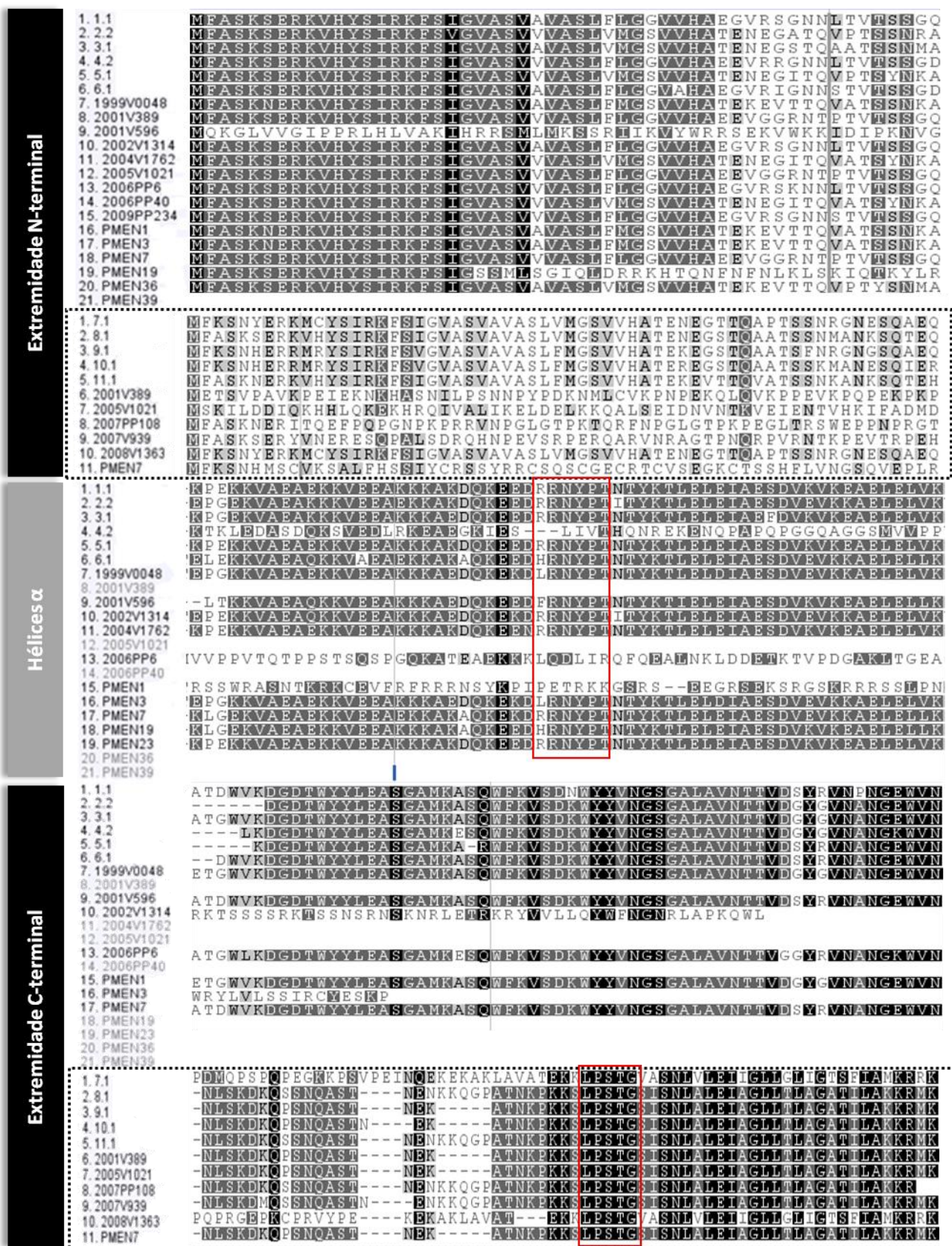


Figura 14: Alinhamento das sequências peptídicas da PspC. Inclui as sequências das PspC já descritas: 1.1, 2.2, 3.1, 4.2, 5.1, 6.1, 7.1, 8.1, 9.1, 10.1 e 11.1, extraídas do GenBank, e alinhadas com as PspC estudadas. As caixas a vermelho evidenciam os motivos encontrados. As caixas a tracejado englobam as PspC com a região de ligação LPxTG.

5. Discussão

5.1 Regiões Variáveis

Streptococcus pneumoniae é uma bactéria naturalmente competente e transformável, tendo por isso uma grande variabilidade genética, resultante de altas taxas de transferência horizontal de genes. Os estudos mais recentes de análises genómicas comparativas de Hakenbeck *et al.* (35) e Tettelin *et al.* (100), determinaram que estirpes individuais de pneumococos variam entre si em mais de 10% do seu conteúdo genómico total. Sandgren *et al.*, concluíram que as propriedades clonais, além de propriedades da própria cápsula, podem ser um fator importante na capacidade do pneumococo causar doença invasiva (87). Assim, esta capacidade é dependente tanto do serotipo como do conteúdo genómico.

Comparações entre serotipos têm o risco de camuflar o impacto de fatores de virulência independentes dos mesmos, devido às diferenças estruturais da cápsula específicas de cada serotipo, que podem afetar a virulência por exemplo ao interferir com a deposição do complemento e com a resistência à fagocitose (51, 66, 67, 108). Consequentemente, no presente estudo procedeu-se à análise dos genomas de estirpes de serotipos 1 e 3, serotipos frequentemente associados à doença invasiva, mas com diferenças significativas entre si (1), de modo a identificar fatores independentes e dependentes do serotipo que alterem a gravidade da doença. Enquanto o serotipo 1 compreende clones com um potencial invasivo homogéneo, o serotipo 3 possui clones com potenciais para causar doença invasiva distintos (86). Para tal, caracterizou-se uma coleção de 540 estirpes de pneumococos, isoladas de infeção invasiva e não invasiva, recuperados em Portugal durante o período de 1999 a 2009 e cinco clones descritos na coleção PMEN (PMEN27, 29, 30, 31 e 40) e procedeu-se à análise da distribuição de oito RDs, identificadas por Harvey *et al.* como sendo importantes para a virulência de *Streptococcus pneumoniae* (41). Este estudo realizou ensaios de virulência em estirpes de serotipo 1 do ST304, ST227, ST3079 e ST3018 e concluiu que as estirpes do ST304 pertencentes à linhagem A são não invasivas, as estirpes do ST227 pertencentes à linhagem A são intermedicamente virulentas e as estirpes dos ST3079 e ST3018 pertencentes às linhagens B e C, correspondentemente, possuem um alto grau de virulência. As RDs analisadas encontraram-se presentes apenas nas estirpes com um alto grau de virulência (41).

O serotipo 1 é considerado como um dos serotipos mais prevalentes isolados de doença invasiva em todo o mundo, tanto em crianças como em adultos, possuindo uma taxa de ataque invulgarmente alta. A sua baixa taxa de colonização pode levar a linhagens geograficamente distintas, que não são encontradas nos outros serotipos (18). Sabe-se que o movimento das pessoas entre países ou continentes transfere frequentemente clones dos serotipos mais comuns, no entanto espera-se que este efeito seja menos ativo no serotipo 1, uma vez que é menos encontrado em colonização. De facto, existe uma estrutura geográfica distinta entre a maioria das linhagens de serotipo 1 (18), originando as 3 linhagens conhecidas: A, B e C. Os

dados obtidos para a presença das RDs nas estirpes recolhidas em Portugal do serotipo 1 apresentam uma grande homogeneidade entre si, uma vez que a maioria das estirpes obteve o mesmo resultado para os *loci* estudados. Esta homogeneidade poderá resultar da estabilidade do serotipo 1, maior do que a dos restantes serotipos, dado que a diversificação do genoma de clones pneumocócicos ocorre principalmente pela acumulação de pequenas sequências importadas de outros pneumococos, o que requer colonização concomitante de estirpes de genótipos distintos (18). Nenhum dos *loci* estudados esteve presente na maioria das estirpes de CC306, exceto o gene *glyco_avir*, que se pensa estar presente apenas nas estirpes avirulentas. A ausência destas RDs nas nossas estirpes pode ser explicada por as estirpes recolhidas em Portugal pertencerem ao CC306, correspondente à linhagem A considerada por Harvey *et al.* como sendo não invasiva ou intermediamente invasiva (41). Contrariamente, as estirpes PMEN27 e 29, que pertencem aos CC217 e CC2296, já obtiveram resultados diferentes. A estirpe PMEN29 apresenta todos os genes exceto o transportador do tipo ABC-2 da RD7 e *glyco_avir*, enquanto a estirpe PMEN27 apresenta a RD1, apesar de não ter o gene codificante da endolisina, possui os genes da família MerR, *rgg/gadr/mutR*, *3HIBDH* e *glyco_avir*. A estirpe PMEN27 pertence à linhagem B e a PMEN29 é do ST615 e pertence à linhagem C, ambas as linhagens foram consideradas por Harvey *et al.*, como sendo altamente virulentas, o que pode explicar a presença da maioria dos *loci* nestas duas estirpes e em nenhuma das estirpes recolhidas em Portugal. Porém pela análise da presença das RDs entre estas duas estirpes observa-se que a PMEN29 possui mais regiões que a PMEN27, parecendo assim que a linhagem C possui mais genes de virulência que a linhagem B. Deste modo, observa-se que as estirpes da linhagem A não possuem nenhuma RD identificada por Harvey *et al.*, a estirpe da linhagem B possui parte das regiões e a linhagem C apresenta a maioria das RDs. Contudo para uma melhor caracterização das linhagens B e C, terá que se estudar um maior número de estirpes pertencentes a estas linhagens do serotipo 1. Outros estudos realizados sobre a presença de RDs em estirpes de diferentes serotipos identificaram diferenças significativas entre estirpes de serotipo 1, pertencentes aos CC306 e CC217 (10, 26). Num destes estudos a diferença entre os clones ocorreu em mais de 40 genes pertencentes a RDs (10).

A caracterização de Harvey *et al.*, que consideram as estirpes da linhagem A como sendo não invasivas, não está em concordância com as informações do comportamento do serotipo 1 em Portugal. Porém estudos anteriores já tinham referido que o serotipo 1 contabiliza tradicionalmente uma maior proporção de infeções invasivas pediátricas na Europa do que na América do Norte (43), apesar das doenças causadas por este serotipo terem sido associadas ao ST306 e STs relacionados, tanto em Portugal e na Europa como na América do Norte (18, 92). Foi sugerido que a percentagem relativamente baixa de IPD causadas pelo serotipo 1 na América do Norte pode refletir o número relativamente elevado de culturas de sangue de bacteriemias ocultas recolhidas nesta zona. É proposto que o serotipo 1 raramente causa bacteriemias ocultas em crianças com menos de dois anos, mas que tende a ser encontrado em pacientes que são

hospitalizados com patologias mais graves. Como consequência, a percentagem de todas as estirpes de IPD causadas por este serotipo na América do Norte parece baixa quando comparada com a da Europa, porque a última compreende na sua maioria amostras recolhidas de doentes hospitalizados (43). Apesar de Harvey *et al.* terem considerado as estirpes de linhagem A como não invasivas, esta corresponde a todo o universo de estirpes de serotipo 1 existente em Portugal. Deste modo, visto que o serotipo 1 é considerado um importante causador de doença invasiva em Portugal, com uma taxa de ataque invulgarmente alta, outras RDs não analisadas neste trabalho deverão estar presentes nestas estirpes promovendo a sua virulência.

O serotipo 3 é dos serotipos mais proeminentes em infeções invasivas em adultos e idosos (97), contrariamente ao que se passa em crianças em que é responsável apenas por 1 a 2% da IPD em todas as regiões geográficas (42). A razão por que este serotipo está relativamente ausente da doença invasiva pediátrica, mas é responsável por uma taxa elevada de doenças em adultos, ainda não é clara. Os resultados obtidos para o serotipo 3 sobre a presença e/ou ausência dos *loci* analisados demonstrou que a maioria das estirpes do CC180 possui os genes de virulência pertencentes à RD1, enquanto os CC260, CC378 e CC458 não possuem estes genes. Enquanto nas RDs 3/4 e 6 a maioria das estirpes do CC180 não apresenta os genes destas regiões, por oposição ao que acontece nas estirpes dos CC260, CC378 e CC458 que apresentam estes genes. O que mais se destaca é as diferenças encontradas entre os clones, contrariamente ao que aconteceu com o serotipo 1. Contudo, visto que este serotipo é considerado um serotipo com clones com potenciais de doença invasiva distintos estas diferenças eram expectáveis. Resultados semelhantes foram encontrados noutros estudos (26, 63). MacAllister *et al.* concluíram que algumas regiões são propriedade de alguns clones enquanto outras são propriedade de todas as estirpes do serotipo 3. Neste estudo pode-se observar que o gene *merR*, gene codificante de um regulador transcricional, esteve presente em 96% das estirpes analisadas (63). É provável que diferentes RDs contribuam de forma diferente para a virulência de estirpes não relacionadas e que a virulência de uma dada estirpe seja uma função de múltiplas RDs. O clone que mais se distinguiu dos restantes foi o CC180. Este clone já foi descrito como sendo um clone dominante, principalmente entre as estirpes de infeção invasiva do serotipo 3 (45), e foi associado aos casos de maior taxa de mortalidade (43%) de todos os clones estudados no trabalho de Sjostrom *et al.* (97).

Encontrou-se também associações entre as estirpes invasivas e as regiões 1, 2 e o sistema *pezAT*. Como já foi referido, o sistema *pezAT* é um sistema toxina-antitoxina (TA) codificado na ilha de patogenicidade *PPI-1*, e este sistema já foi implicado na virulência (15). Foram propostas diferentes funções para os sistemas TA como controlo de crescimento, estabilização de elementos genéticos móveis ou controlo da morte celular programada (61). As estirpes pneumocócicas contendo *PPI-1* codificam vários fatores de virulência que são muito vantajosos para a infeção e possuem uma maior patogenicidade. O gene *pezA* funciona como um repressor que se liga a uma sequência a montante do mesmo e o gene *pezT* serve como co-repressor do complexo *pezA-pezT*, de forma a continuar a reprimir *pezA*. Estirpes sem o gene *pezT* possuem um progresso

de infecção mais lento (15). Por outro lado, as estirpes não invasivas apresentaram uma associação com o gene *merR* e com a RD7. É de realçar que existem associações com as regiões RD1, RD2 e RD7, mas não com todos os seus genes constituintes. O caso que mais se destaca é a associação entre RD2 e as estirpes invasivas, contrária à associação do gene *merR*, localizado nesta região, com as estirpes não invasivas. Apesar desta associação o sistema MerR parece ser um mecanismo importante para a bactéria se proteger de níveis elevados de NO durante o processo infeccioso, mas parece não ser importante durante a fase de colonização da nasofaringe e tem um efeito modesto nos pulmões (98). Contudo visto que este gene esteve presente na maioria das estirpes analisadas é provável que além do papel conhecido para a infecção invasiva, seja também importante para outros tipos de infecção.

Foram encontradas também diferenças entre a distribuição dos genes dentro da mesma região. Por exemplo, a maioria das estirpes do CC180 possui o sistema toxina-antitoxina *pezAT*, mas não possui os genes *rgg/gadR/mutR* e *3HIBDH* pertencentes a esta região. Estes resultados sugerem mais uma vez que além da presença das RDs poder ser muito variável entre estirpes, a própria composição das regiões é muito inconstante. Porém determinou-se uma associação entre a presença dos genes *int* e *pblB*, *rgg/gadR/mutR* e Fe^{2+}/Pb^{2+} e *ABC* e Fe^{2+}/Pb^{2+} . Ou seja, existe uma alta probabilidade de estes pares de genes estarem ambos presentes ou ausentes no genoma da mesma estirpe. Os genes Fe^{2+}/Pb^{2+} e *ABC* também tiveram uma forte associação com o tipo de CC, isto é, a presença e a ausência de ambos os genes está relacionada com o tipo de CC a que uma dada estirpe pertence. Este par de genes parece estar ausente no CC180 e presente nos CC260, CC378 e CC447. O impacto de qualquer combinação de RDs pode ser influenciado por outros fatores, incluindo a redundância das funções codificadas nas RDs e serotipo capsular, assim como fatores do hospedeiro (63).

Concluiu-se que, para os três maiores clones analisados, a presença da RD1 completa aparenta ser uma propriedade clonal das estirpes invasivas de CC180, CC260 e CC378 e a RD2 completa, uma propriedade das estirpes invasivas de CC180 e CC260. Por outro lado o gene MerR (RD2) parece ser uma propriedade clonal das estirpes não invasivas novamente do CC180, os genes *pezAT* relacionaram-se com as estirpes invasivas do CC180 e também do CC260 e a RD7 é uma propriedade das estirpes invasivas dos CC260, CC378 e CC458 e uma propriedade das estirpes não invasivas do CC180. Em síntese o CC180 parece ter como propriedade clonal para as estirpes invasivas a presença de RD1 e RD2 e os genes *pezAT* e *rgg/gadR/mutR* de RD3/4 e para as estirpes não invasivas os genes *merR* e a RD7. Para os CC260 e CC378 a RD1 e RD7 parecem ser propriedades clonais das estirpes invasivas, assim como a RD2 mas apenas para as estirpes do CC260. Apenas a RD7 parece ser propriedade tanto das estirpes invasivas como não invasivas, os restantes *loci* relacionaram-se apenas com um dos tipos de infecção. Estes resultados justificam-se por os genes das RDs poderem codificar para funções diretamente associadas à virulência da bactéria como adesão, invasão, toxicidade, e evasão ao sistema imunitário do hospedeiro ou contribuírem para funções metabólicas

necessárias para o crescimento em diferentes nichos do hospedeiro (10). Havendo, assim regiões mais necessárias para a invasão e outras mais necessárias para a colonização.

Globalmente parece não haver uma resposta simples para a pergunta sobre a prevalência de uma estirpe em doença invasiva em vez de em colonização. Enquanto os polissacarídeos capsulares são absolutamente necessários para a virulência, a relação do serotipo e do genoma com o grau de virulência é complexa e ainda não está totalmente esclarecida. Foi sugerido que a capacidade de cada serotipo causar doença invasiva é dependente de um conjunto distinto de fatores genéticos (75). Outra hipótese indicada é que as RDs podem contribuir para atenuar a virulência (62), ou que a influência do tipo de polissacarídeo capsular no potencial da estirpe para causar doença invasiva é tão esmagadora que a influência de outros fatores é silenciada (86). Ao compararmos os resultados dos dois serotipos sugerimos que as RDs são necessárias para completar as propriedades fisiológicas únicas de cada tipo capsular, sendo por isso diferentes para cada serotipo e dentro deles para cada CC. A própria constituição das regiões parece ser diferente entre CCs. Sandgren *et al.* concluíram que os serotipos pneumocócicos dividem-se em 2 grupos, com classes que se sobrepõem. O primeiro grupo é constituído pelos serotipos mais invasivos, como o serotipo 1 (incluem também os serotipos 4, 7F e 9V) que estão ausentes ou são raramente encontrados na colonização e por isso são clonalmente relacionados. Estes clones são claramente agentes patogénicos eficientes que ao entrarem no hospedeiro têm a capacidade de causar doença. Consequentemente, possuem possivelmente atributos específicos que lhes permitem ser mais invasivos e/ou mais capazes de manipular o sistema imunitário do hospedeiro. O outro grupo é constituído pelos serotipos 6A, 6B, 14 e 19F que, como o serotipo 3, são encontrados em colonização e em doença invasiva. Estirpes pertencentes a estes serotipos têm uma maior diversidade genética, especialmente as que são encontradas na nasofaringe (87). Deste modo, a capacidade de uma estirpe causar doença invasiva parece depender de um conjunto de fatores que englobam o serotipo, genótipos e fatores do hospedeiro (63), havendo algumas RDs que parecem ser necessárias para causar doença invasiva enquanto outras são necessárias para a colonização.

5.2 Proteína de Superfície Pneumocócica C

A PspC é uma proteína de superfície exposta, multifuncional e um importante fator de virulência com um papel ativo na invasão e patogenicidade deste microrganismo.

O *locus pspC* foi caracterizado para um número significativo de estirpes causadoras de doença e foi encontrada uma grande variabilidade ao nível da sequência deste gene (12, 52). Apesar desta variabilidade, as proteínas PspC têm uma organização comum, exceto na região C-terminal, apresentando na região N-terminal hélices α expostas na superfície bacteriana, seguidas de uma região rica em prolinas (12), que aglomera cinco domínios denominados de A, B, R1, R2 e C (60, 112). Iannelli *et al.* propuseram uma nova classificação que visa distinguir a variedade da PspC, organizando-a em 11 grupos diferentes de acordo com

a sua estrutura e função (52), e em duas grandes famílias, de acordo com o motivo de ligação da região C-terminal. A primeira família inclui as PspC dos grupos 1 a 6, que possuem uma região com repetições do motivo de ligação à colina na região C-terminal, que liga a proteína à superfície bacteriana através de interações com a colina presente nos ácidos teicóicos e lipoteicóicos. Enquanto a segunda inclui os grupos 7 a 11 que têm um motivo de ligação LPxTG típico de bactérias de Gram-positivo. Devido a esta multiplicidade de fenótipos, é importante investigar a variabilidade do *locus pspC* dentro de diferentes estirpes, de modo a adquirir uma melhor compreensão da PspC como fator de virulência e facilitar o desenho de vacinas baseadas nesta proteína.

Neste trabalho caracterizou-se o *locus pspC* de 19 estirpes pneumocócicas, de diferentes serotipos e de diferentes tipos infeção. Também se procedeu à análise das sequências peptídicas da PspC depositadas no GenBank.

5.2.1 Estrutura da PspC

Começou-se por realizar uma análise conformacional às sequências peptídicas de PspC depositadas no GenBank, semelhante à análise realizada em outros estudos (38, 52, 60). Deste modo, pôde-se verificar que além da sequência sinal que é idêntica em todos os grupos, o domínio A é o mais conservado dos domínios, nomeadamente dentro da família das proteínas com ligação colina, com exceção das proteínas do grupo 5 que possuem domínios com um maior número de aminoácidos. As diferenças encontradas no grupo 5 podem ser explicadas por este grupo apresentar uma maior semelhança com a proteína PspA, sendo esta normalmente uma proteína maior que a PspC (12, 52).

Em geral o grupo da PspC a que uma dada estirpe pertence é independente do serotipo, uma vez que estirpes do mesmo serotipo apresentam PspCs de diferentes grupos, porém já foram encontradas algumas correlações. Aparentemente as estirpes de serotipo 3 apresentam apenas PspCs dos grupos 8 a 11 (52), esta correlação também foi encontrada neste trabalho, onde as duas estirpes estudadas de serotipo 3 expressam proteínas do grupo 8 e 11. Foi encontrada também uma correlação entre as PspCs do grupo 4 e os serotipos 4, 6B, 7F e principalmente 14 (29). As duas estirpes correspondentes a PspC4 analisadas neste trabalho são dos serotipos 14 e 19A, contudo a estirpe do serotipo 19A apresentou uma baixa identidade com as proteínas do grupo 4. As estirpes encontradas com duas cópias da PspC englobaram os grupos 4-10, 6-7 e 6-9. A associação dos grupos 4 -10 e 6-9 foi descrita por outros autores (52), contudo a conjugação dos grupos 6 e 7 não tinha sido ainda reportada.

Ao analisar-se globalmente todas as sequências peptídicas da PspC verificou-se uma grande variabilidade dentro dos grupos, principalmente nos grupos 1 a 6, os grupos com motivos de ligação à colina. É de salientar as diferenças encontradas dentro do grupo 3, que apresenta proteínas de diferentes tamanhos e conformações. Por exemplo, foi encontrada uma grande semelhança entre a PspC3.4 e a PspC2.2, com uma identidade de 99,7%, superior à identidade encontrada entre a PspCs do grupo 3. Devido a estes erros nas

classificações e ao facto de não ser, por vezes, muito claro a diferenciação entre alguns grupos seria conveniente adotar-se parâmetros mais rígidos para a separação das PspCs em grupos ou talvez, devido à grande variabilidade das PspCs, seja impossível a sua classificação em grupos. Possivelmente seria preferível distingui-los apenas em duas grandes famílias a de ligação LPxTG e a de ligação à colina ou, em alternativa, proceder a estudos funcionais da proteína que ajudassem a uma mais clara classificação.

5.2.2 Funções da PspC

A PspC é uma proteína de superfície exposta que se liga a diversas proteínas do hospedeiro como: IgA secretória, o recetor de imunoglobulina polimérica (pIgR) e os fatores H, C3 e C4BP do complemento. Um importante componente do sistema imunitário do hospedeiro é o sistema do complemento, constituído aproximadamente por 30 proteínas solúveis e proteínas da superfície celular organizadas em três cascatas enzimáticas denominadas de via clássica, via alternativa e via da lectina. Tanto a via clássica como a alternativa levam à formação de C3-convertase que cliva o componente C3, fator central do complemento, resultando na deposição de C3b na superfície do agente patogénico que é processada em iC3b.

Até à data a PspC é a única proteína pneumocócica que se liga ao fator H e estudos demonstraram que esta ligação parece ser independente do serotipo e do tipo de PspC (27). Para confirmar estas afirmações realizou-se a metodologia “Far-Western Blot” às 19 estirpes estudadas, pertencentes a diferentes serotipos e que se determinou possuírem PspC de diferentes grupos, de modo a distinguir as estirpes que se ligavam ao fator H. Todas as estirpes analisadas apresentaram ligação ao fator H, o que é congruente com os resultados de outros trabalhos (27, 28, 80). O tamanho das bandas variou entre as estirpes e não houve nenhuma correlação entre o serotipo capsular ou tipo de PspC com a ligação ao fator H, uma vez que todas as estirpes se ligaram a este fator. Estes resultados são concordantes com outro trabalho que analisou 18 estirpes por “Western Blot”, pertencentes a diferentes serotipos (27). Contudo outros estudos analisaram a força da ligação da PspC ao fator H por citometria de fluxo, e concluíram que o grau de ligação ao fator H varia entre estirpes e é dependente da PspC, mas esta interação é afetada pelo serotipo capsular (111). A eficiência da ligação ao fator H aumenta significativamente quando são usadas estirpes não capsuladas, indicando que a capsula polissacarídica interfere com a ligação ao fator H (36). Recentemente também foi detetada a ligação da PspC ao fator do complemento C4BP (29). Esta ligação parece ser dependente de uma ou mais propriedades genéticas presentes em tipos de clones específicos do pneumococo ou ser influenciada pela presença de outras proteínas do soro (29).

Outra função da PspC é mediar a aderência pneumocócica através da ligação ao recetor da imunoglobulina polimérica (pIgR), pelo domínio extracelular “Ig-like”, também conhecido como componente secretório (“Secretory Component”, SC). O SC liga-se covalentemente a dímeros da imunoglobulina IgA que são transportados, por um mecanismo chamado de transcitose¹, da superfície basolateral para a superfície

¹ Transporte de moléculas de um lado ao outro do epitélio por endocitose, através de vesículas.

apical das células da nasofaringe epitelial, onde o complexo SC/IgA₂ é clivado em plgR de modo a libertar IgA secretória (slgA). Quando o pneumococo invade as células do epitélio da nasofaringe do hospedeiro a PspC liga-se especificamente ao domínio SC de plgR e despoleta a maquinaria de modo a translocar, por transcitose, o pneumococo até à corrente sanguínea (112). Estudos recentes demonstraram que a ligação entre a PspC e o domínio SC ocorre através de um motivo constituído por um hexapéptido que está localizado nos domínios R1 e R2, da PspC (60). Análises densitométricas revelaram porém, que apenas a sequência YPT é crítica para esta ligação (39). A análise das sequências peptídicas das nossas estirpes revelou que 84,2% das estirpes possuem hexapéptidos com motivos de ligação ao domínio SC e que estes motivos variaram no primeiro aminoácido (^{F,R,Y,L}/_{RNYPT}), mas têm sempre em comum o motivo RNYPT. Destas estirpes, apenas 25% apresentaram dois motivos, um em cada domínio R. Apenas é necessário um motivo num dos domínios R para ocorrer a ligação a SC de plgR, ou a SC como parte de slgA (28). Os dados aqui presentes mostram que as PspC do grupo 3, 4, 6 e 7 só possuem um destes motivos no domínio R1 tal como foi descrito por Iannelli *et al.* e Moreno *et al.* (52, 71), e que as PspC do grupo 7 são o único grupo da família de ligação LPxTG que apresenta este domínio (52). Neste trabalho observou-se uma variação da presença dos domínios e na sua localização, mesmo dentro dos grupos. Pôde-se observar que as PspCs pertencentes ao grupo 3 e 6 apresentaram os dois domínios, apenas um domínio em R1, apenas um domínio em R2 ou a presença de nenhum domínio. A PspC do grupo 7 apresentou apenas um motivo YPT no domínio R1. Este facto demonstra, mais uma vez, a grande variabilidade dentro de grupos, principalmente nos grupos da família de ligação à colina e que, por conseguinte a ligação a SC parece não estar relacionada com o tipo de PspC. Atendendo a estes resultados conclui-se que, aparentemente, a PspC usa dois epítomos diferentes para a ligação dos fatores do hospedeiro H e SC e estudos independentes revelaram que estas ligações são independentes entre si (36).

Apesar dos fatores que podem afetar as interações da PspC ao complemento serem provavelmente muito complexos, resultados indicam que o conteúdo genético (“background”) da estirpe deve ser o fator com maior influência nos efeitos de perda de deposição de C3b/iC3b, independentemente do serotipo capsular (111). Tendo em conta estes dados pode-se extrapolar estas conclusões para a ligação da PspC aos restantes fatores do complemento. Os efeitos do conteúdo genético da estirpe podem ser mediados pela extensa variação na estrutura alélica da PspC (52) e talvez por outras causas de variação não capsular entre estirpes de *S. pneumoniae*, como a variação alélica de PspA ou de outras proteínas de superfície, a presença ou ausência de regiões de diversidade, ou múltiplas deleções e inserções únicas de genes (100).

5.3 Limitações e perspectivas futuras do estudo

Este estudo apresenta algumas limitações: 1) o número reduzido de estirpes de alguns STs limita algumas conclusões. 2) A presença das regiões completas é identificada, por PCR, pela ausência de amplificação do fragmento de DNA. Estes dados não diferenciam a existência da região completa da existência de apenas parte da região. 3) O desenho dos oligonucleotídeos iniciadores nos genes flanqueantes leva apenas à conclusão que determinado *locus* está presente/ausente naquela zona do genoma, podendo-se encontrar noutro local do genoma. 4) A confirmação dos dados de PCR pela técnica “Dot Blot” ainda não se encontra concluída, não tendo por isso sido apresentados os resultados neste trabalho. Estes dados darão uma maior fiabilidade aos resultados obtidos.

Como perspectivas de continuação de trabalho futuras propõe-se terminar os resultados de hibridação por “Dot Blot”, de modo a dar uma maior robustez aos resultados obtidos por PCR. Além disso, seria pertinente realizar um estudo semelhante com as mesmas estirpes noutras RDs, descritas por outros autores. As RDs analisadas neste trabalho foram identificadas em estirpes de serotipo 1, por isso seria interessante analisar a presença de outras RDs identificadas em estirpes de outros serotipos e observar o seu comportamento na nossa coleção.

Para estudar o fator de virulência PspC pensamos que seria interessante realizar mutantes, em estirpes acapsuladas, com as regiões C-terminais trocadas. Ou seja, começaríamos por realizar um “knock-out” ao gene *pspC* e estudaríamos o seu comportamento em modelos de infeção em ratinhos. De seguida, transformávamos uma estirpe com uma PspC da família de ligação à colina (por exemplo a estirpe TIGR4, que possui uma PspC3.4), retirando a sua região C-terminal com motivos de ligação à colina e colocando no seu lugar uma região de ligação LPxTG. A mesma metodologia seria realizada numa estirpe da família de ligação LPxTG, colocando uma região com motivos de ligação à colina. O comportamento dos mutantes seria depois observado em modelos de infeção em ratinhos. Deste modo, esperaríamos conseguir concluir qual a importância dos dois motivos de ligação, e talvez perceber a razão da existência das duas regiões C-terminais na mesma proteína.

6. Conclusões

No presente trabalho caracterizou-se a presença de oito RDs com um potencial na virulência de *S. pneumoniae* numa coleção de 540 estirpes de pneumococos recuperados em Portugal (de 1999 a 2009), de serotipo 1 e 3 e de cinco estirpes da coleção PMEN. Os dados obtidos para a presença das RDs nas estirpes de serotipo 1 permitiu confirmar a grande homogeneidade existente em serotipos raramente presentes em colonização. Determinou-se que apesar desta homogeneidade, o serotipo 1 distribui-se em 3 linhagens distintas geograficamente que apresentam conjuntos de RDs diferentes. Deste modo, conclui-se que as estirpes da linhagem A não possuem nenhuma RD estudada, a estirpe da linhagem B possui todas as RDs estudadas exceto o gene codificante da endolisina (*lyt*), a RD2 e a RD7 e a estirpe da linhagem C apresenta todas as RDs exceto o gene *lyt* e RD7. Os resultados obtidos para o serotipo 3 demonstraram que o clone que mais se distinguiu foi o CC180, um clone dominante principalmente em estirpes de infeção invasiva. Quanto à presença dos *loci* a maioria das estirpes do CC180 possui os genes de virulência pertencentes à RD1 e não apresenta a maioria das restantes RDs. Enquanto os CC260, CC378 e CC458 não possuem os genes da RD1, mas apresentam os genes das RDs 3/4 e 6. Foram encontradas também diferenças entre a distribuição dos genes dentro da mesma região. O que sugere que, além do número e do conjunto de RDs poder ser muito variável entre estirpes, a própria composição das regiões é muito inconstante. Porém determinou-se uma associação quanto à presença de genes pertencentes à mesma RD, assim como a RDs distintas. Os genes codificantes para a permease de alta afinidade com $\text{Fe}^{2+}/\text{Pb}^{2+}$ e o transportador do tipo ABC-2 possuem uma forte associação com o tipo de CC, isto é, a presença e a ausência de ambos os genes está relacionada com o tipo de CC a que uma dada estirpe pertence. Este par de genes parece estar ausente no CC180 e presente nos CC260, CC378 e CC447. Demonstrou-se que alguns *loci* aparentam ser propriedades clonais de alguns CC, relacionado com um dos tipos de doença, e outros, como a RD7, relacionam-se com os dois tipos de doença, o que sugere que os genes das RDs podem codificar para funções diretamente associadas à virulência da bactéria da mesma forma que podem contribuir para funções metabólicas necessárias para o crescimento em diferentes nichos do hospedeiro. Existem, assim regiões mais necessárias para a invasão e outras mais necessárias para a colonização do pneumococo, refletindo o comportamento dos serotipos 1 e 3, que normalmente apresentam associações diferentes com o hospedeiro. Deste modo, as RDs são necessárias para completar as propriedades fisiológicas únicas de cada tipo capsular, sendo por isso diferentes para cada serotipo e dentro deles para cada CC. A diversidade dos próprios fatores de virulência, como a PspC, parece ser muito variável entre estirpes e nem sempre relacionada com o serotipo ou com o CC das estirpes. Os efeitos do conteúdo genético no grau de virulência de uma dada estirpe é função de um conjunto de fatores que englobam a conjugação e distribuição de RDs, a variação alélica da PspC, assim como de outros fatores de virulência, que em conjunto com o serotipo capsular, originam diferenças entre a patogenicidade das estirpes pneumocócicas.

7. Anexos

Quadro 14: Genes constituintes das RDs e sua localização em genomas já sequenciados.

Região	Anotação/Função	ORF ID das estirpes anotadas de <i>Streptococcus pneumoniae</i>		
		P1031 (SPP)	TIGR4 (SP)	G54 (SPG)
RD1	<i>Integrase</i> (Int)	0028	-	-
	<i>Proteína PblB-like</i> (PblB)	0075	-	-
	<i>Amidase lítica</i> ("N-acetylmuramoyl-L-alanine amidase", lyt)	0083	-	-
RD2	<i>Regulador transcricional da família MerR</i> (MerR)	0750	0737	0669
RD3	<i>PezA</i> (antitoxina)	1053	1050	-
	<i>PezT</i> (toxina)	1054	1051	-
RD4	<i>Regulador transcricional da família rgg/gadR/mutR</i> (Rgg)	1060	-	0975
	<i>3HIBDH</i>	1062	-	0977
RD5	<i>ZmpD</i>	1141	-	-
RD6	<i>Permease de alta afinidade com Fe²⁺/Pb²⁺</i> (Fe ²⁺ /Pb ²⁺)	1340	1300	1194
RD7	<i>Transportador do tipo ABC-2</i> (ABC)	1779	-	-
RD8	<i>Proteína da família hidrolase glicosídica</i> (Glyco)	1353	-	-

Adaptado de (41)

Quadro 15: Coleção de estirpes para a análise da PspC

Estirpes	Serotipo	ST	Tipo de Infecção	Local de recolha
1999V0048	14	156	Invasiva	Sangue
2001V0389	19A	193	Invasiva	Sangue
2001V0596	8	53	Invasiva	Sangue
2002V1314	6A	460	Não Invasiva	Expetoração
2004V1762	19A	276	Não Invasiva	Expetoração
2005V1021	6C	1715	Invasiva	Sangue
2006PP0006	14	9	Não Invasiva	Pus
2006PP0040	19A	276	Invasiva	LCR
2007PP108	3	232	Não Invasiva	Pus
2007V0939	3	180	Invasiva	LCR
2008V1363	9V	280	Invasiva	Sangue
2009PP234	7F	4771	Invasiva	Sangue
PMEN1	23F	81	Invasiva	
PMEN 3	9V	156	Invasiva	
PMEN7	19A	75	Invasiva	
PMEN 19	5	289	Invasiva	
PMEN23	6A	376	Invasiva	
PMEN 36	18C	113	Invasiva	
PMEN39	7F	191	Invasiva	

LCR-Líquido Cefalorraquidiano

Quadro 16: Oligonucleotídeos iniciadores utilizado nas reações de PCR

Propósito	Oligonucleotídeo iniciador	Sequência (5'- 3')	Dimensão (bp) ¹
MLST ²	<i>aroE</i> _M13F	tgt aaa acg acg gcc agt CGT TTA GCT GCA GTT GTT GC	405
	<i>aroE</i> _M13R-pUC	cag gaa aca gct atg acC CCC ACA CTG GTG GCA TTA AC	
	<i>gdh</i> _M13F	tgt aaa acg acg gcc agt GTG CTG AAA AGA TTA AGG TCT	460
	<i>gdh</i> _M13R-pUC	cag gaa aca gct atg acC TGC TTC CAG CTT TAT AGT CAT G	
	<i>gki</i> _M13F	tgt aaa acg acg gcc agt GGC ATT GGA ATG GGA TCA CC	483
	<i>gki</i> _M13R-pUC	cag gaa aca gct atg acC TCT CCC GCA GCT GAC AC	
	<i>recP</i> _M13F	tgt aaa acg acg gcc agt GCC AAC TCA GGT CAT CCA GG	450
	<i>recP</i> _M13R-pUC	cag gaa aca gct atg acC GCT TCC AAG TCT GTT CCA TTT TC	
	<i>spi</i> _M13F	tgt aaa acg acg gcc agt CGC TTA GAA AGG TAA GTT ATG	474
	<i>spi</i> _M13R-pUC	cag gaa aca gct atg acC AGG CTG AGA TTG GTG ATT CTC	
	<i>xpt</i> _M13F	tgt aaa acg acg gcc agt GGA GGT CTT ATG AAA TTA TTA G	486
	<i>xpt</i> _M13R-pUC	cag gaa aca gct atg acC AGA TCT GCC TCC TTA AAT AC	
RDs	<i>ddl</i> _M13F	tgt aaa acg acg gcc agt TTG CCA TGG ATA AAA TCA CGA C	441
	<i>ddl</i> _M13R-pUC	cag gaa aca gct atg acC CGC GTT GTA AAA CTT TCC	
	RD1_fw	GGTCTGGTCGTGAACAAAC	304
	RD1_rv	GCCTCTCTCAAAGCCTCCCTC	
	PblB_fw	GAGGAGAGCACGAGACAAGCGA	581 ³
	PblB_rv	CGAGCAACCTTATCCGCAATT	
	Int_rv	AAAGGCAAGCGTTACTATGTC	1377 ^{3,4}
	Lyt_fw	CGGTTGGAATGCTGAGAGTTAT	1376 ^{3,5}
	RD2_fw2	GTTCACCCAGACGATGCTTATG	3563
	RD2_rv	GCCCTCTGTACCGGCTGGG	
	RD2_fw	GAGGAAGAGGACGTAGAAAATGG	1749
	MerR_rv	CAGTCAAATCCATAAGTGCG	
	PezA_fw	GGGTGCGGCAAGGGATTTT	2621
	PezT_rv	TCTGCTGCACTAATTTCTACTCCATG	
	Rgg/GadR/MutR_fw	AGTAACCGGTTGCATAGAGAAA	707 ⁶
	Rgg/GadR/MutR_rv	ACTTGGCATCACACTTAGAAA	
	3-HIBDH_fw	CCTGAAACAGTCGAAAAAAT	753 ⁶
	3-HIBDH_rv	AAGCAATCCCTATATTGCTG	
	ZmpD_fw	GGGTATTGTTGGTTCGGTATT	462 ⁶
	ZmpD_rv	GATGCTGGAGAAACCGACGA	
	RpmE2_fw	GTTCACTCAAGCAGATGGACG	488 ⁷
	Fe/Pb_rv	ATAAGGATTCTGAGGTTGGTAAG	
	RD7_fw	CATAGACTTCCACCAGACAGG	1594
	RD7_rv	GACTATTGACGATGCCTATGAT	
	ABC_fw	GGAAGCACCGTTATCAAGAG	596 ³
	ABC_rv	GTTGGTGCTTATTTGACGC	
	Glyco_vir_fw	CCTTCGATTGCGTAGATTTTAGC	721 ³
	Glyco_vir_rv	GGAGAGATTTCATTGCGCTTGC	
	Gyco_avir_fw	GATAGCGGTTTAGTAGTTGGGTA	230 ⁸
	Glyco_avir_rv	CATCATCGTAAGTTTGAAGCT	

¹ Dimensão do fragmento em TIGR4.

² As letras maiúsculas das sequências iniciadores usadas em MLST correspondem a sequências que hibridam com o genoma de pneumococos, as letras minúsculas referem-se aos iniciadores M13F e M13R-pUC, que foram adicionadas para facilitar a sequenciação.

³ Dimensão do fragmento em P1031.

⁴ Dimensão do fragmento com RD1_fw

⁵ Dimensão do fragmento com RD1_rv

⁶ Gene ausente em TIGR4, dimensão em G54.

⁷ Gene ausente em TIGR4 e G54, dimensão em R6.

⁸ Controlo Hungary-19A

Quadro 17: Valores de “odds ratio” (OR) e de Fisher (valores p) referentes ao serotipo 1.

RDs	306 (n=61)		217(n=1)		2296 (n=1)	
	OR	Valores p	OR	Valores p	OR	Valores p
RD1	0,000	0,001*	Inf	0,032*	Inf	0,032*
PbIB	0,000	0,018	Inf	0,143	Inf	0,143
Int	0,000	0,001*	Inf	0,032*	Inf	0,032*
Lyt	Inf	1,000	0,000	1,000	0,000	1,000
RD2	0,023	0,063	0,000	1,000	Inf	0,032
MerR	0,000	0,001*	Inf	0,032*	Inf	0,032*
pezAT	0,000	0,032*	0,000	1,000	Inf	0,016*
Rgg	0,000	0,001*	Inf	0,032*	Inf	0,032*
3HIBDH	0,000	0,001*	Inf	0,032*	Inf	0,032*
ZmpD	0,042	0,095	0,000	1,000	Inf	0,048
Fe/Pb	0,000	0,032*	0,000	1,000	Inf	0,016*
RD7	0,000	0,033*	0,000	1,000	Inf	0,033*
ABC	1,000	1,000	0,000	1,000	0,000	1,000
vir	0,024	0,064	0,000	1,000	Inf	0,032
avir	12,633	0,156	Inf	1,000	0,000	0,081

OR>1 indica uma associação entre a presença dos *loci* e o CC correspondente, OR≤1 indica uma associação entre a ausência dos *loci* e o CC correspondente. Os valores a vermelho representam os valores estatisticamente significativos pelo teste de Fisher (p≤0,05). * Indica que é estatisticamente significativo pelo teste de correção de falsos positivos (FDR≤0,05). “Inf”-Infinito. *Int*-integrase, *Lyt*-endolisina, *vir*-família hidrólase glicosídica do tipo virulento, *avir*- família hidrólase glicosídica do tipo avirulento, *Rgg*- regulador transcricional RGG/GadR/MutR.

Quadro 18: Valores de “odds ratio” (OR) e de Fisher (valores p) referentes ao serotipo 3.

RDs	62 (n=5)		63 (n=1)		180 (n=284)		260 (n=98)		378 (n=73)		447 (n=4)		458 (n=15)		1087 (n=1)	
	OR	Valores p	OR	Valores p	OR	Valores p	OR	Valores p	OR	Valores p	OR	Valores p	OR	Valores p	OR	Valores p
RD1	0,906	1,000	0,000	1,000	0,763	0,260	2,091	0,004*	0,927	0,878	0,000	0,582	0,000	0,050	0,000	1,000
PbIB	0,593	0,671	Inf	1,000	19,411	≤0,001*	0,091	≤0,001*	0,135	≤0,001*	0,893	1,000	0,000	≤0,001*	Inf	1,000
Int	0,656	0,684	Inf	1,000	36,760	≤0,001*	0,057	≤0,001*	0,053	≤0,001*	0,000	0,060	0,067	0,000*	Inf	1,000
Lyt	0,920	1,000	0,000	1,000	21,314	≤0,001*	0,100	≤0,001*	0,062	≤0,001*	0,000	0,143	0,000	0,000*	Inf	0,420
RD2	0,872	1,000	Inf	1,000	0,816	0,336	1,497	0,102	0,864	0,599	0,000	0,018	2,380	0,276	Inf	1,000
MerR	Inf	1,000	Inf	1,000	0,372	0,095	2,226	0,389	1,542	0,752	Inf	1,000	Inf	1,000	Inf	1,000
pezAT	0,092	0,015	0,000	0,125	0,956	0,890	1,027	1,000	1,702	0,335	0,139	0,078	Inf	0,235	Inf	1,000
Rgg	2,575	0,364	Inf	0,370	0,035	≤0,001*	11,855	≤0,001*	14,101	≤0,001*	Inf	0,018*	0,000	0,002*	Inf	0,370
3HIBDH	157,533	≤0,001*	Inf	0,015	0,273	0,128	0,000	0,354	0,931	1,000	0,000	1,000	0,000	1,000	0,000	1,000
ZmpD	107,955	≤0,001*	Inf	0,019*	0,341	0,170	0,000	0,215	0,695	1,000	0,000	1,000	0,000	1,000	Inf	0,019*
Fe/Pb	0,000	0,167	0,000	1,000	0,014	≤0,001*	25,675	≤0,001*	27,716	≤0,001*	5,600	0,127	0,000	0,002*	0,000	1,000
RD7	1,333	1,000	0,000	0,470	1,766	≤0,001*	0,776	0,307	0,692	0,162	0,000	0,048	0,433	0,187	0,000	0,470
ABC	2,598	0,362	0,000	1,000	0,014	≤0,001*	32,827	≤0,001*	18,438	≤0,001*	Inf	0,018*	0,000	0,002*	Inf	0,368
vir	47,678	≤0,001*	0,000	1,000	0,473	0,139	0,512	0,544	2,421	0,157	0,000	1,000	0,000	1,000	0,000	1,000
avir	0,000	1,000	0,000	1,000	0,805	0,778	0,000	0,080	1,703	0,428	130,810	≤0,001*	0,000	1,000	0,000	1,000

OR>1 indica uma associação entre a presença dos *loci* e o CC correspondente, OR≤1 indica uma associação entre a ausência dos *loci* e o CC correspondente. Os valores a vermelho representam os valores estatisticamente significativos pelo teste de Fisher (p≤0,05). * Indica que é estatisticamente significativo pelo teste de correção de falsos positivos (FDR≤0,05). “Inf”-Infinito. *Int*-integrase, *Lyt*-endolisina, *vir*-família hidrólase glicosídica do tipo virulento, *avir*- família hidrólase glicosídica do tipo avirulento, *Rgg*- regulador transcricional RGG/GadR/MutR.

8. Referências

1. **Aguiar, S. I., M. J. Brito, J. Goncalo-Marques, J. Melo-Cristino, and M. Ramirez.** 2010. Serotypes 1, 7F and 19A became the leading causes of pediatric invasive pneumococcal infections in Portugal after 7 years of heptavalent conjugate vaccine use. *Vaccine* **28**:5167-73.
2. **Aguiar, S. I., M. J. Frias, L. Santos, J. Melo-Cristino, and M. Ramirez.** 2006. Emergence of optochin resistance among *Streptococcus pneumoniae* in Portugal. *Microb Drug Resist* **12**:239-45.
3. **Avery, O. T., and R. Dubos.** 1931. The Protective Action of a Specific Enzyme against Type Iii *Pneumococcus* Infection in Mice. *J Exp Med* **54**:73-89.
4. **Bagos, P. G., T. D. Liakopoulos, I. C. Spyropoulos, and S. J. Hamodrakas.** 2004. PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acids Res* **32**:W400-4.
5. **Behr, T., W. Fischer, J. Peter-Katalinic, and H. Egge.** 1992. The structure of pneumococcal lipoteichoic acid. Improved preparation, chemical and mass spectrometric studies. *Eur J Biochem* **207**:1063-75.
6. **Benin, A. L., K. L. O'Brien, J. P. Watt, R. Reid, E. R. Zell, S. Katz, C. Donaldson, A. Parkinson, A. Schuchat, M. Santosham, and C. G. Whitney.** 2003. Effectiveness of the 23-valent polysaccharide vaccine against invasive pneumococcal disease in Navajo adults. *J Infect Dis* **188**:81-9.
7. **Benjamini, Y., D. Drai, G. Elmer, N. Kafkafi, and I. Golani.** 2001. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125**:279-84.
8. **Berry, A. M., and J. C. Paton.** 2000. Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infect Immun* **68**:133-40.
9. **Berry, A. M., J. Yother, D. E. Briles, D. Hansman, and J. C. Paton.** 1989. Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. *Infect Immun* **57**:2037-42.
10. **Blomberg, C., J. Dagerhamn, S. Dahlberg, S. Browall, J. Fernebro, B. Albiger, E. Morfeldt, S. Normark, and B. Henriques-Normark.** 2009. Pattern of accessory regions and invasive disease potential in *Streptococcus pneumoniae*. *J Infect Dis* **199**:1032-42.
11. **Bogaert, D., R. De Groot, and P. W. Hermans.** 2004. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis* **4**:144-54.
12. **Brooks-Walter, A., D. E. Briles, and S. K. Hollingshead.** 1999. The *pspC* gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. *Infect Immun* **67**:6533-42.
13. **Brown, J. S., S. M. Gilliland, and D. W. Holden.** 2001. A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. *Mol Microbiol* **40**:572-85.
14. **Brown, J. S., S. M. Gilliland, J. Ruiz-Albert, and D. W. Holden.** 2002. Characterization of *pit*, a *Streptococcus pneumoniae* iron uptake ABC transporter. *Infect Immun* **70**:4389-98.
15. **Brown, J. S., S. M. Gilliland, B. G. Spratt, and D. W. Holden.** 2004. A locus contained within a variable region of pneumococcal pathogenicity island 1 contributes to virulence in mice. *Infect Immun* **72**:1587-93.
16. **Bruckner, R., M. Nuhn, P. Reichmann, B. Weber, and R. Hakenbeck.** 2004. Mosaic genes and mosaic chromosomes-genomic variation in *Streptococcus pneumoniae*. *Int J Med Microbiol* **294**:157-68.
17. **Brueggemann, A. B., T. E. Peto, D. W. Crook, J. C. Butler, K. G. Kristinsson, and B. G. Spratt.** 2004. Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. *J Infect Dis* **190**:1203-11.
18. **Brueggemann, A. B., and B. G. Spratt.** 2003. Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. *J Clin Microbiol* **41**:4966-70.
19. **Busenlehner, L. S., M. A. Pennella, and D. P. Giedroc.** 2003. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiol Rev* **27**:131-43.
20. **Calix, J. J., and M. H. Nahm.** 2010. A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. *J Infect Dis* **202**:29-38.
21. **Camilli, R., E. Pettini, M. Del Grosso, G. Pozzi, A. Pantosti, and M. R. Oggioni.** 2006. Zinc metalloproteinase genes in clinical isolates of *Streptococcus pneumoniae*: association of the full array with a clonal cluster comprising serotypes 8 and 11A. *Microbiology* **152**:313-21.
22. **Canvin, J. R., A. P. Marvin, M. Sivakumaran, J. C. Paton, G. J. Boulnois, P. W. Andrew, and T. J. Mitchell.** 1995. The role of pneumolysin and autolysin in the pathology of pneumonia and septicemia in mice infected with a type 2 pneumococcus. *J Infect Dis* **172**:119-23.
23. **Carriço, J. A., F. R. Pinto, C. Simas, S. Nunes, N. G. Sousa, N. Frazao, H. de Lencastre, and J. S. Almeida.** 2005. Assessment of band-based similarity coefficients for automatic type and subtype classification of microbial isolates analyzed by pulsed-field gel electrophoresis. *J Clin Microbiol* **43**:5483-90.
24. **Carriço, J. A., C. Silva-Costa, J. Melo-Cristino, F. R. Pinto, H. de Lencastre, J. S. Almeida, and M. Ramirez.** 2006. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *J Clin Microbiol* **44**:2524-32.
25. **Cheng, Q., D. Finkel, and M. K. Hostetter.** 2000. Novel purification scheme and functions for a C3-binding protein from *Streptococcus pneumoniae*. *Biochemistry* **39**:5450-7.

26. **Dagerhamn, J., C. Blomberg, S. Browall, K. Sjöström, E. Morfeldt, and B. Henriques-Normark.** 2008. Determination of accessory gene patterns predicts the same relatedness among strains of *Streptococcus pneumoniae* as sequencing of housekeeping genes does and represents a novel approach in molecular epidemiology. *J Clin Microbiol* **46**:863-8.
27. **Dave, S., A. Brooks-Walter, M. K. Pangburn, and L. S. McDaniel.** 2001. PspC, a pneumococcal surface protein, binds human factor H. *Infect Immun* **69**:3435-7.
28. **Dave, S., S. Carmicle, S. Hammerschmidt, M. K. Pangburn, and L. S. McDaniel.** 2004. Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H. *J Immunol* **173**:471-7.
29. **Dieudonne-Vatran, A., S. Krentz, A. M. Blom, S. Meri, B. Henriques-Normark, K. Riesbeck, and B. Albiger.** 2009. Clinical isolates of *Streptococcus pneumoniae* bind the complement inhibitor C4b-binding protein in a PspC allele-dependent fashion. *J Immunol* **182**:7865-77.
30. **Donati, C., N. L. Hiller, H. Tettelin, A. Muzzi, N. J. Croucher, S. V. Angiuoli, M. Oggioni, J. C. Dunning Hotopp, F. Z. Hu, D. R. Riley, A. Covacci, T. J. Mitchell, S. D. Bentley, M. Kilian, G. D. Ehrlich, R. Rappuoli, E. R. Moxon, and V. Masignani.** 2010. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol* **11**:R107.
31. **Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* **186**:1518-30.
32. **Fenoll, A., J. J. Granizo, L. Aguilar, M. J. Gimenez, L. Aragonese-Fenoll, G. Hanquet, J. Casal, and D. Tarrago.** 2009. Temporal trends of invasive *Streptococcus pneumoniae* serotypes and antimicrobial resistance patterns in Spain from 1979 to 2007. *J Clin Microbiol* **47**:1012-20.
33. **Francisco, A. P., M. Bugalho, M. Ramirez, and J. A. Carrico.** 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* **10**:152.
34. **Garnier, J., J. F. Gibrat, and B. Robson.** 1996. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol* **266**:540-53.
35. **Hakenbeck, R., N. Balmelle, B. Weber, C. Gardes, W. Keck, and A. de Saizieu.** 2001. Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. *Infect Immun* **69**:2477-86.
36. **Hammerschmidt, S., V. Agarwal, A. Kunert, S. Haelbich, C. Skerka, and P. F. Zipfel.** 2007. The host immune regulator factor H interacts via two contact sites with the PspC protein of *Streptococcus pneumoniae* and mediates adhesion to host epithelial cells. *J Immunol* **178**:5848-58.
37. **Hammerschmidt, S., G. Bethe, P. H. Remane, and G. S. Chhatwal.** 1999. Identification of pneumococcal surface protein A as a lactoferrin-binding protein of *Streptococcus pneumoniae*. *Infect Immun* **67**:1683-7.
38. **Hammerschmidt, S., S. R. Talay, P. Brandtzaeg, and G. S. Chhatwal.** 1997. SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol Microbiol* **25**:1113-24.
39. **Hammerschmidt, S., M. P. Tillig, S. Wolff, J. P. Vaerman, and G. S. Chhatwal.** 2000. Species-specific binding of human secretory component to SpsA protein of *Streptococcus pneumoniae* via a hexapeptide motif. *Mol Microbiol* **36**:726-36.
40. **Harboe, Z. B., T. L. Benfield, P. Valentiner-Branth, T. Hjuler, L. Lambertsen, M. Kaltoft, K. Krogh, H. C. Slotved, J. J. Christensen, and H. B. Konradsen.** 2010. Temporal trends in invasive pneumococcal disease and pneumococcal serotypes over 7 decades. *Clin Infect Dis* **50**:329-37.
41. **Harvey, R. M., U. H. Stroehrer, A. D. Ogunniyi, H. C. Smith-Vaughan, A. J. Leach, and J. C. Paton.** 2011. A variable region within the genome of *Streptococcus pneumoniae* contributes to strain-strain variation in virulence. *PLoS One* **6**:e19650.
42. **Hausdorff, W. P., J. Bryant, P. R. Paradiso, and G. R. Siber.** 2000. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* **30**:100-21.
43. **Hausdorff, W. P., G. Siber, and P. R. Paradiso.** 2001. Geographical differences in invasive pneumococcal disease rates and serotype frequency in young children. *Lancet* **357**:950-2.
44. **Hava, D. L., and A. Camilli.** 2002. Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Mol Microbiol* **45**:1389-406.
45. **Henriques, B., M. Kalin, A. Ortqvist, B. Olsson Liljequist, M. Almela, T. J. Marrie, M. A. Mufson, A. Torres, M. A. Woodhead, S. B. Svenson, and G. Kallenius.** 2000. Molecular epidemiology of *Streptococcus pneumoniae* causing invasive disease in 5 countries. *J Infect Dis* **182**:833-9.
46. **Hiller, N. L., A. Ahmed, E. Powell, D. P. Martin, R. Eutsey, J. Earl, B. Janto, R. J. Boissy, J. Hogg, K. Barbadora, R. Sampath, S. Lonergan, J. C. Post, F. Z. Hu, and G. D. Ehrlich.** 2010. Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog* **6**:e1001108.
47. **Hiller, N. L., B. Janto, J. S. Hogg, R. Boissy, S. Yu, E. Powell, R. Keefe, N. E. Ehrlich, K. Shen, J. Hayes, K. Barbadora, W. Klimke, D. Dernovoy, T. Tatusova, J. Parkhill, S. D. Bentley, J. C. Post, G. D. Ehrlich, and F. Z. Hu.** 2007. Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. *J Bacteriol* **189**:8186-95.
48. **Hollands, A., R. K. Aziz, R. Kansal, M. Kotb, V. Nizet, and M. J. Walker.** 2008. A naturally occurring mutation in *ropB* suppresses SpeB expression and reduces M1T1 group A streptococcal systemic virulence. *PLoS One* **3**:e4102.
49. **Horácio, A. N., J. Diamantino-Miranda, S. Aguiar, A. H. Ramirez, J. Melo-Cristino, and P. G. f. t. S. o. S. Infections.** 2013. The Majority of Adult Pneumococcal Invasive Infections in Portugal Are Still Potentially Vaccine Preventable in Spite of Significant Declines of Serotypes 1 and 5. *PLoS One* **8**:e73704.
50. **Horácio, A. N., J. Diamantino-Miranda, S. I. Aguiar, M. Ramirez, and J. Melo-Cristino.** 2012. Serotype changes in adult invasive pneumococcal infections in Portugal did not reduce the high fraction of potentially vaccine preventable infections. *Vaccine* **30**:218-24.

51. Hyams, C., J. Yuste, K. Bax, E. Camberlein, J. N. Weiser, and J. S. Brown. 2010. Streptococcus pneumoniae resistance to complement-mediated immunity is dependent on the capsular serotype. *Infect Immun* **78**:716-25.
52. Iannelli, F., M. R. Oggioni, and G. Pozzi. 2002. Allelic variation in the highly polymorphic locus *pspC* of Streptococcus pneumoniae. *Gene* **284**:63-71.
53. Janulczyk, R., F. Iannelli, A. G. Sjöholm, G. Pozzi, and L. Björck. 2000. Hic, a novel surface protein of Streptococcus pneumoniae that interferes with complement function. *J Biol Chem* **275**:37257-63.
54. Jedrzejewski, M. J. 2001. Pneumococcal virulence factors: structure and function. *Microbiol Mol Biol Rev* **65**:187-207 ; first page, table of contents.
55. Kadioglu, A., J. N. Weiser, J. C. Paton, and P. W. Andrew. 2008. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol* **6**:288-301.
56. Kerr, A. R., G. K. Paterson, J. McCluskey, F. Iannelli, M. R. Oggioni, G. Pozzi, and T. J. Mitchell. 2006. The contribution of PspC to pneumococcal virulence varies between strains and is accomplished by both complement evasion and complement-independent mechanisms. *Infect Immun* **74**:5319-24.
57. Kilian, M., K. Poulsen, T. Blomqvist, L. S. Havarstein, M. Bek-Thomsen, H. Tettelin, and U. B. Sørensen. 2008. Evolution of Streptococcus pneumoniae and its close commensal relatives. *PLoS One* **3**:e2683.
58. Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**:1759-64.
59. Lau, G. W., S. Haataja, M. Lonetto, S. E. Kensit, A. Marra, A. P. Bryant, D. McDevitt, D. A. Morrison, and D. W. Holden. 2001. A functional genomic analysis of type 3 Streptococcus pneumoniae virulence. *Mol Microbiol* **40**:555-71.
60. Luo, R., B. Mann, W. S. Lewis, A. Rowe, R. Heath, M. L. Stewart, A. E. Hamburger, S. Sivakolundu, E. R. Lacy, P. J. Björkman, E. Tuomanen, and R. W. Kriwacki. 2005. Solution structure of choline binding protein A, the major adhesin of Streptococcus pneumoniae. *EMBO J* **24**:34-43.
61. Magnuson, R. D. 2007. Hypothetical functions of toxin-antitoxin systems. *J Bacteriol* **189**:6089-92.
62. Maurelli, A. T., R. E. Fernandez, C. A. Bloch, C. K. Rode, and A. Fasano. 1998. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli. *Proc Natl Acad Sci U S A* **95**:3943-8.
63. McAllister, L. J., A. D. Ogunniyi, U. H. Stroehner, A. J. Leach, and J. C. Paton. 2011. Contribution of serotype and genetic background to virulence of serotype 3 and serogroup 11 pneumococcal isolates. *Infect Immun* **79**:4839-49.
64. McDaniel, L. S., J. Yother, M. Vijayakumar, L. McGarry, W. R. Guild, and D. E. Briles. 1987. Use of insertional inactivation to facilitate studies of biological properties of pneumococcal surface protein A (PspA). *J Exp Med* **165**:381-94.
65. McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefevre, A. Tomasz, and K. P. Klugman. 2001. Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. *J Clin Microbiol* **39**:2565-71.
66. Melin, M., K. Trzcinski, M. Antonio, S. Meri, R. Adegbola, T. Kaijalainen, H. Kayhty, and M. Vakevainen. 2010. Serotype-related variation in susceptibility to complement deposition and opsonophagocytosis among clinical isolates of Streptococcus pneumoniae. *Infect Immun* **78**:5252-61.
67. Melin, M., K. Trzcinski, S. Meri, H. Kayhty, and M. Vakevainen. 2010. The capsular serotype of Streptococcus pneumoniae is more important than the genetic background for resistance to complement. *Infect Immun* **78**:5262-70.
68. Melo-Cristino, J., L. Santos, and M. Ramirez. 2006. The Viriato Study: update of antimicrobial susceptibility data of bacterial pathogens from community-acquired respiratory tract infections in Portugal in 2003 and 2004. *Rev Port Pneumol* **12**:9-30.
69. Mitchell, J., I. R. Siboo, D. Takamatsu, H. F. Chambers, and P. M. Sullam. 2007. Mechanism of cell surface expression of the Streptococcus mitis platelet binding proteins PblA and PblB. *Mol Microbiol* **64**:844-57.
70. Mitchell, T. J., P. W. Andrew, F. K. Saunders, A. N. Smith, and G. J. Boulnois. 1991. Complement activation and antibody binding by pneumolysin via a region of the toxin homologous to a human acute-phase protein. *Mol Microbiol* **5**:1883-8.
71. Moreno, A. T., M. L. Oliveira, P. L. Ho, C. F. Vadesilho, G. M. Palma, J. M. Ferreira, Jr., D. M. Ferreira, S. R. Santos, M. B. Martinez, and E. N. Miyaji. 2012. Cross-reactivity of antipneumococcal surface protein C (PspC) antibodies with different strains and evaluation of inhibition of human complement factor H and secretory IgA binding via PspC. *Clin Vaccine Immunol* **19**:499-507.
72. Musher, D. M. 1992. Infections caused by Streptococcus pneumoniae: clinical spectrum, pathogenesis, immunity, and treatment. *Clin Infect Dis* **14**:801-7.
73. Nelson, A. L., A. M. Roche, J. M. Gould, K. Chim, A. J. Ratner, and J. N. Weiser. 2007. Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infect Immun* **75**:83-90.
74. O'Brien, K. L., L. J. Wolfson, J. P. Watt, E. Henkle, M. Deloria-Knoll, N. McCall, E. Lee, K. Mulholland, O. S. Levine, and T. Cherian. 2009. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet* **374**:893-902.
75. Obert, C., J. Sublett, D. Kaushal, E. Hinojosa, T. Barton, E. I. Tuomanen, and C. J. Orihuela. 2006. Identification of a Candidate Streptococcus pneumoniae core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect Immun* **74**:4766-77.
76. Orihuela, C. J., J. N. Radin, J. E. Sublett, G. Gao, D. Kaushal, and E. I. Tuomanen. 2004. Microarray analysis of pneumococcal gene expression during invasive disease. *Infect Immun* **72**:5582-96.
77. Paton, J. C., B. Rowan-Kelly, and A. Ferrante. 1984. Activation of human complement by the pneumococcal toxin pneumolysin. *Infect Immun* **43**:1085-7.

78. **Polissi, A., A. Pontiggia, G. Feger, M. Altieri, H. Mottl, L. Ferrari, and D. Simon.** 1998. Large-scale identification of virulence genes from *Streptococcus pneumoniae*. *Infect Immun* **66**:5620-9.
79. **Poolman, J., P. Kriz, C. Feron, E. Di-Paolo, I. Henckaerts, A. Miseur, D. Wauters, R. Prymula, and L. Schuerman.** 2009. Pneumococcal serotype 3 otitis media, limited effect of polysaccharide conjugate immunisation and strain characteristics. *Vaccine* **27**:3213-22.
80. **Quin, L. R., C. Onwubiko, S. Carmicle, and L. S. McDaniel.** 2006. Interaction of clinical isolates of *Streptococcus pneumoniae* with human complement factor H. *FEMS Microbiol Lett* **264**:98-103.
81. **Regev-Yochay, G., M. Raz, R. Dagan, N. Porat, B. Shainberg, E. Pinco, N. Keller, and E. Rubinstein.** 2004. Nasopharyngeal carriage of *Streptococcus pneumoniae* by adults and children in community and family settings. *Clin Infect Dis* **38**:632-9.
82. **Ren, B., A. J. Szalai, O. Thomas, S. K. Hollingshead, and D. E. Briles.** 2003. Both family 1 and family 2 PspA proteins can inhibit complement deposition and confer virulence to a capsular serotype 3 strain of *Streptococcus pneumoniae*. *Infect Immun* **71**:75-85.
83. **Rodrigues, F., S. Nunes, R. Sa-Leao, G. Goncalves, L. Lemos, and H. de Lencastre.** 2009. *Streptococcus pneumoniae* nasopharyngeal carriage in children attending day-care centers in the central region of Portugal, in the era of 7-valent pneumococcal conjugate vaccine. *Microb Drug Resist* **15**:269-77.
84. **Rosen, J. B., A. R. Thomas, C. A. Lexau, A. Reingold, J. L. Hadler, L. H. Harrison, N. M. Bennett, W. Schaffner, M. M. Farley, B. W. Beall, and M. R. Moore.** 2011. Geographic variation in invasive pneumococcal disease following pneumococcal conjugate vaccine introduction in the United States. *Clin Infect Dis* **53**:137-43.
85. **Rosenow, C., P. Ryan, J. N. Weiser, S. Johnson, P. Fontan, A. Ortqvist, and H. R. Masure.** 1997. Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Mol Microbiol* **25**:819-29.
86. **Sá-Leao, R., F. Pinto, S. Aguiar, S. Nunes, J. A. Carrico, N. Frazao, N. Goncalves-Sousa, J. Melo-Cristino, H. de Lencastre, and M. Ramirez.** 2011. Analysis of invasiveness of pneumococcal serotypes and clones circulating in Portugal before widespread use of conjugate vaccines reveals heterogeneous behavior of clones expressing the same serotype. *J Clin Microbiol* **49**:1369-75.
87. **Sandgren, A., K. Sjostrom, B. Olsson-Liljequist, B. Christensson, A. Samuelsson, G. Kronvall, and B. Henriques Normark.** 2004. Effect of clonal and serotype-specific properties on the invasive capacity of *Streptococcus pneumoniae*. *J Infect Dis* **189**:785-96.
88. **Sato, K., M. K. Quartey, C. L. Liebler, C. T. Le, and G. S. Giebink.** 1996. Roles of autolysin and pneumolysin in middle ear inflammation caused by a type 3 *Streptococcus pneumoniae* strain in the chinchilla otitis media model. *Infect Immun* **64**:1140-5.
89. **Schuerman, L., D. Borys, B. Hoet, A. Forsgren, and R. Prymula.** 2009. Prevention of otitis media: now a reality? *Vaccine* **27**:5748-54.
90. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**:67-75.
91. **Senkovich, O., W. J. Cook, S. Mirza, S. K. Hollingshead, Protasevich, II, D. E. Briles, and D. Chattopadhyay.** 2007. Structure of a complex of human lactoferrin N-lobe with pneumococcal surface protein a provides insight into microbial defense mechanism. *J Mol Biol* **370**:701-13.
92. **Serrano, I., J. Melo-Cristino, J. A. Carrico, and M. Ramirez.** 2005. Characterization of the genetic lineages responsible for pneumococcal invasive disease in Portugal. *J Clin Microbiol* **43**:1706-15.
93. **Serrano, I., M. Ramirez, and J. Melo-Cristino.** 2004. Invasive *Streptococcus pneumoniae* from Portugal: implications for vaccination and antimicrobial therapy. *Clin Microbiol Infect* **10**:652-6.
94. **Severiano, A., F. R. Pinto, M. Ramirez, and J. A. Carrico.** 2011. Adjusted Wallace coefficient as a measure of congruence between typing methods. *J Clin Microbiol* **49**:3997-4000.
95. **Silva, N. A., J. McCluskey, J. M. Jefferies, J. Hinds, A. Smith, S. C. Clarke, T. J. Mitchell, and G. K. Paterson.** 2006. Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. *Infect Immun* **74**:3513-8.
96. **Simões, A. S., L. Pereira, S. Nunes, A. Brito-Avo, H. de Lencastre, and R. Sa-Leao.** 2011. Clonal evolution leading to maintenance of antibiotic resistance rates among colonizing *Pneumococci* in the PCV7 era in Portugal. *J Clin Microbiol* **49**:2810-7.
97. **Sjostrom, K., C. Spindler, A. Ortqvist, M. Kalin, A. Sandgren, S. Kuhlmann-Berenzon, and B. Henriques-Normark.** 2006. Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. *Clin Infect Dis* **42**:451-9.
98. **Stroeher, U. H., S. P. Kidd, S. L. Stafford, M. P. Jennings, J. C. Paton, and A. G. McEwan.** 2007. A pneumococcal MerR-like regulator and S-nitrosoglutathione reductase are required for systemic virulence. *J Infect Dis* **196**:1820-6.
99. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33**:2233-9.

100. Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**:498-506.
101. Tseng, H. J., A. G. McEwan, J. C. Paton, and M. P. Jennings. 2002. Virulence of *Streptococcus pneumoniae*: PsaA mutants are hypersensitive to oxidative stress. *Infect Immun* **70**:1635-9.
102. Tuomanen, E. 1999. Molecular and cellular biology of pneumococcal infection. *Curr Opin Microbiol* **2**:35-9.
103. Tuomanen, E., H. Liu, B. Hengstler, O. Zak, and A. Tomasz. 1985. The induction of meningeal inflammation by components of the pneumococcal cell wall. *J Infect Dis* **151**:859-68.
104. Tuomanen, E. T., T. J. Mitchell, D. A. Morrison, and B. G. Spratt. 2004. *The pneumococcus*, 1 ed. ASM Press.
105. Van der Poll, T., and S. M. Opal. 2009. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet* **374**:1543-56.
106. Vesikari, T., J. Wysocki, B. Chevallier, A. Karvonen, H. Czajka, J. P. Arsene, P. Lommel, I. Dieussaert, and L. Schuerman. 2009. Immunogenicity of the 10-valent pneumococcal non-typeable *Haemophilus influenzae* protein D conjugate vaccine (PHiD-CV) compared to the licensed 7vCRM vaccine. *Pediatr Infect Dis J* **28**:S66-76.
107. Watson, D. A., and D. M. Musher. 1990. Interruption of capsule production in *Streptococcus pneumoniae* serotype 3 by insertion of transposon Tn916. *Infect Immun* **58**:3135-8.
108. Weinberger, D. M., K. Trzcinski, Y. J. Lu, D. Bogaert, A. Brandes, J. Galagan, P. W. Anderson, R. Malley, and M. Lipsitch. 2009. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. *PLoS Pathog* **5**:e1000476.
109. Yeh, S. H., A. Gurtman, D. C. Hurley, S. L. Block, R. H. Schwartz, S. Patterson, K. U. Jansen, J. Love, W. C. Gruber, E. A. Emini, and D. A. Scott. 2010. Immunogenicity and safety of 13-valent pneumococcal conjugate vaccine in infants and toddlers. *Pediatrics* **126**:e493-505.
110. Yother, J., and J. M. White. 1994. Novel surface attachment mechanism of the *Streptococcus pneumoniae* protein PspA. *J Bacteriol* **176**:2976-85.
111. Yuste, J., S. Khandavilli, N. Ansari, K. Muttardi, L. Ismail, C. Hyams, J. Weiser, T. Mitchell, and J. S. Brown. 2010. The effects of PspC on complement-mediated immunity to *Streptococcus pneumoniae* vary with strain background and capsular serotype. *Infect Immun* **78**:283-92.
112. Zhang, J. R., K. E. Mostov, M. E. Lamm, M. Nanno, S. Shimida, M. Ohwaki, and E. Tuomanen. 2000. The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell* **102**:827-37.